chist-era

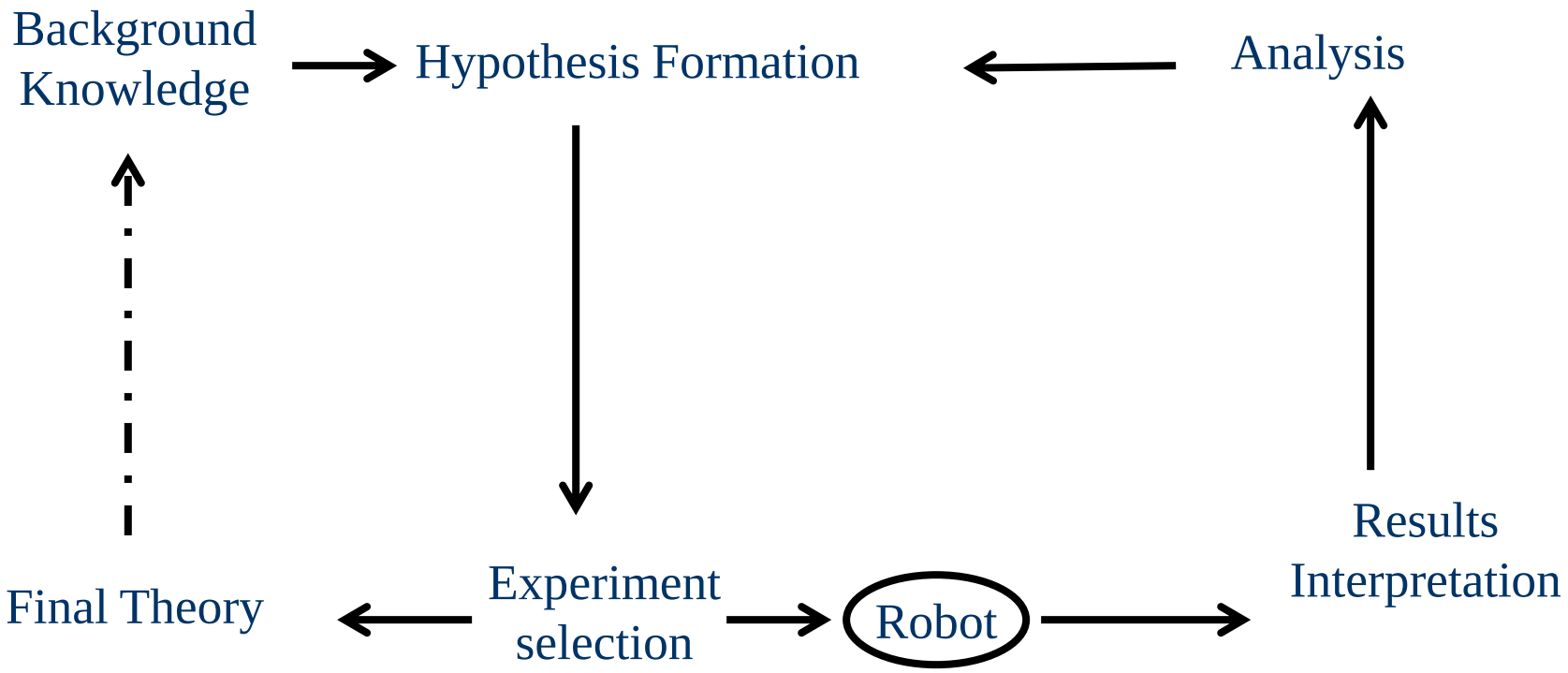
AdaLab

AdaLab

Scientific Background

The Concept of a Robot Scientist

Computer systems capable of originating their own experiments, physically executing them, interpreting the results, and then repeating the cycle.



Eve: Robot Scientist

The Reproducibility Crisis

- One of the most important current issues in biology is ‘The reproducibility crisis’ - Billions of euros wasted.

- ‘There is growing alarm about results that cannot be reproduced. Explanations include increased levels of scrutiny, complexity of experiments and statistics, and pressures on researchers.’ *Nature* 2016.

- *We require Automation to ensure reproducibility.*

Key Goals

Scientific Goals

- To make scientific research more efficient: cheaper, faster, better.

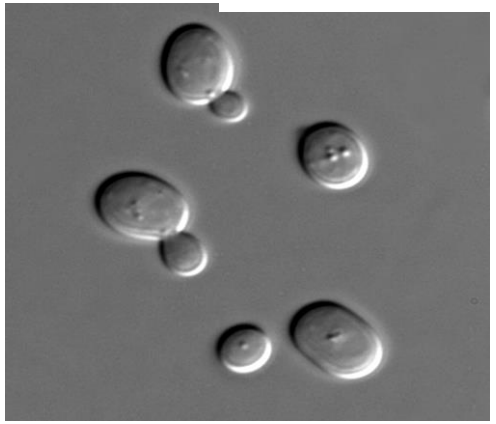
- Our vision is that within 10 years many scientific discoveries will be made by teams of human and robot scientists.
- This collaborations will produce scientific knowledge more efficiently than either could alone.

Scientific Goals

- A framework for semi-automated and automated knowledge discovery by teams of human and robot scientists.

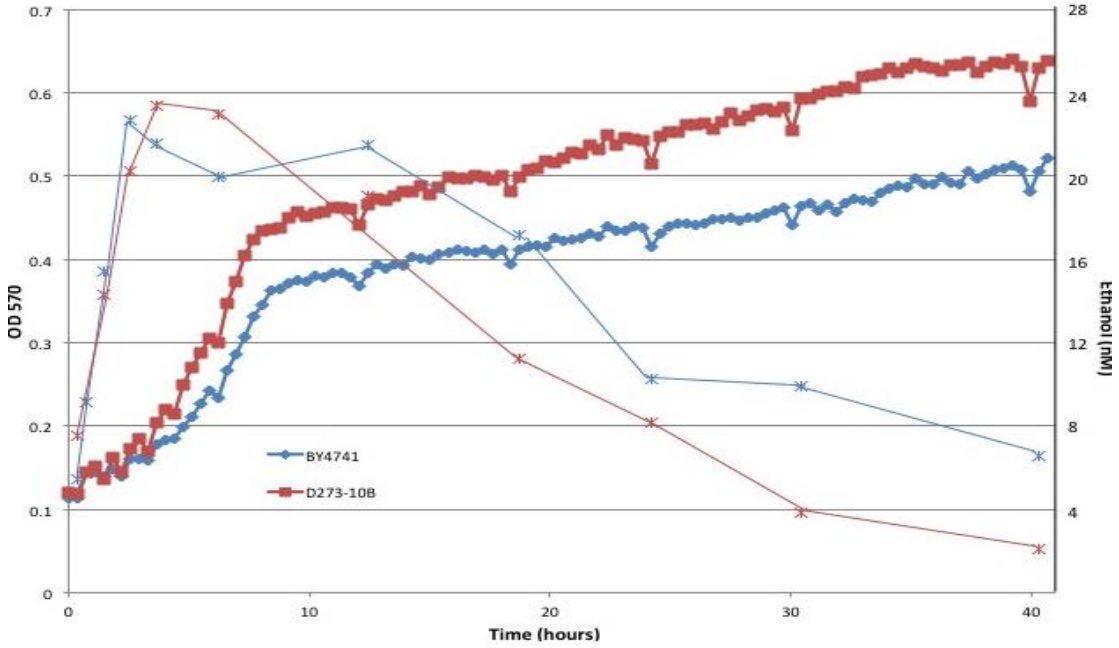
- Integrating advances in knowledge representation, ontology engineering, semantic technologies, machine learning, bioinformatics, and automated experimentation.

The Diauxic Shift



- Yeast (*S. cerevisiae*).
- First turn sugar into ethanol.
- Then turn ethanol into CO_2.

- Cancer
- Ageing



Key Challenges

Key Challenges

- The AdaLab system needs to be:
  - *autonomous and perceptive to human requirements* (its scientific collaborators).
  - *able to continuously learn, adapt and improve in the "real world" complex environment* of scientific research.
  - capable of continuous cycles of scientific hypothesis formation and experimentation that will improve its scientific knowledge (models).

Key Challenges

- Scientific knowledge is inherently uncertain.
- Within the AdaLab framework we are developing Bayesian methods that make inferences and plan experiments under uncertainty.

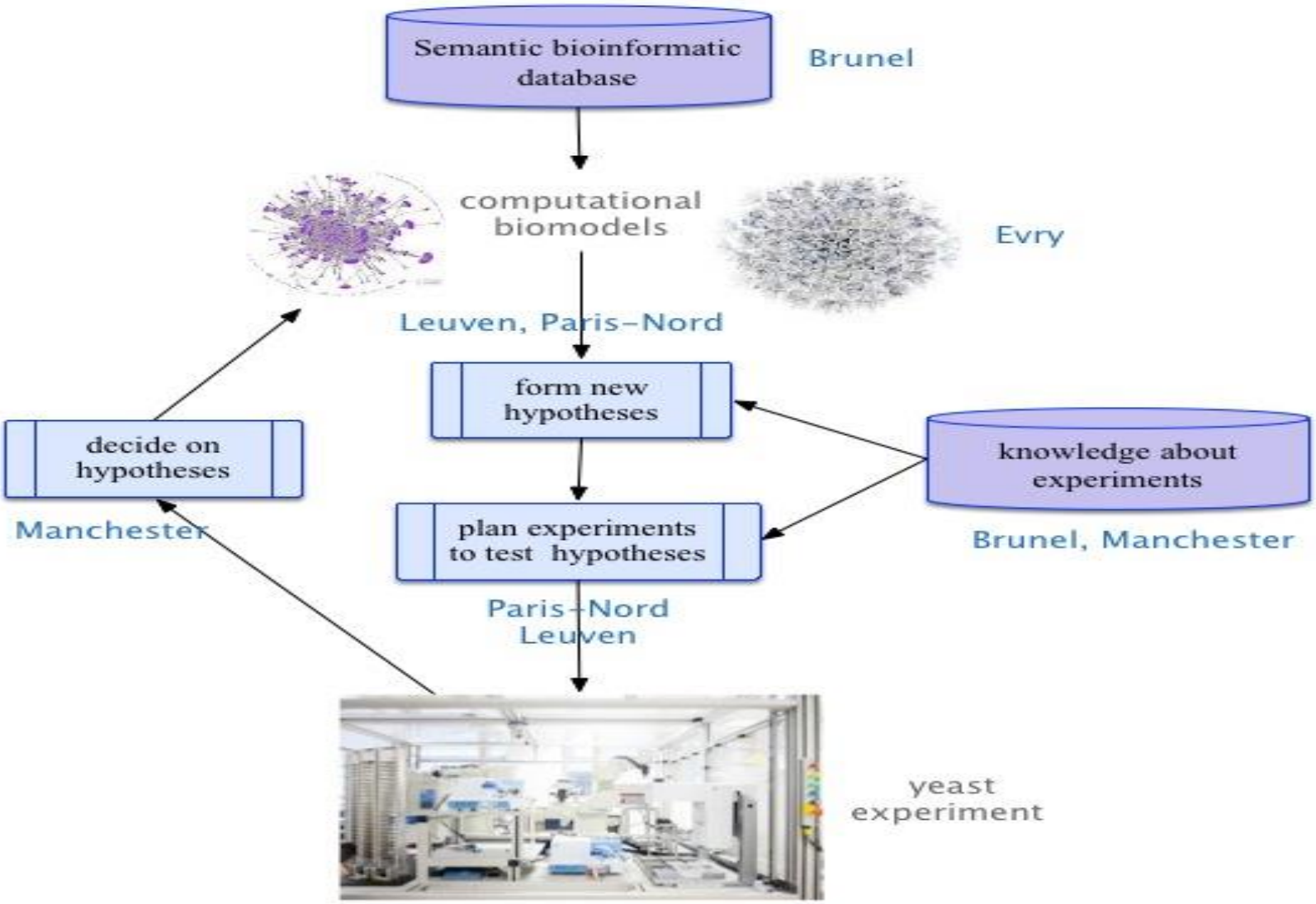
- Scientific knowledge is best represented using logic.
- To integrate logic with probabilities we will use statistical relational learning, and have developed an ontology for representing uncertain knowledge.

chist-era

AdaLab

Structure/Partners

AdaLab Structure



Brunel University

- Coordination
- Knowledge representation – ontologies.
- Formalisation of the knowledge base on the yeast diauxic shift.
- A communication mechanism between robot and human scientists.

Key Outputs - Brunel

Knowledge Representation

- Ontologies:
  - UNO (uncertainties ontology) has been updated and submitted to an open public repository
  - Meta-data ontology (for description of the datasets) and models has been extended
  - AdaLab ontology (models biological entities) has been significantly extended and submitted to an open public repository
- A knowledge base about diauxic shift has been produced
- An intermediate version of a communication mechanism has been developed

KULeuven

- Knowledge Engineering.
- Machine Learning: ILP.
- Probabilistic Inference.

Key Outputs – Leuven

Machine Learning 2

- Learning biological networks from data
  - Partially from literature, partly from new experiments, Learn partial models. Objective measure: simulator
- Designing experiments
  - Find experiments most informative for improving biological network model. ng most uncertain outcome.
- Integrating better priors and more domain knowledge..

Universite Paris-Nord

- Machine Learning
  - Inductive Logic Programming.
  - Incremental theory revision.
- Experiment planning
  - Active learning.
  - Cost minimisation.

Key Outputs – Paris Nord

Machine Learning

- Learning probabilistic graphical models from scarce data :
  - Learn new regulatory links
  - general method : good results on DREAM D8C1 challenges 3rd from ~100.
- Learning attributed graph patterns
  - method designed to handle structural/frequency-based constraints.
  - Mining co-regulation graphs – investigating biology of patterns.

Universite Evry

- Bioinformatics: Collection of bioinformatic data about the yeast diauxic shift.
- Systems Biology: Development of an integrated metabolic and gene signalling network.

Key Outputs – Evry

Systems Biology Modelling

- Statistical learning methods for joint inference of a metabolic-regulatory model.
- A version of the simulation tools, including both regulatory and metabolic model simulation, has been finalized.
- Phenotype predictions from genotype using the integrated model.

University of Manchester: Robot Scientists

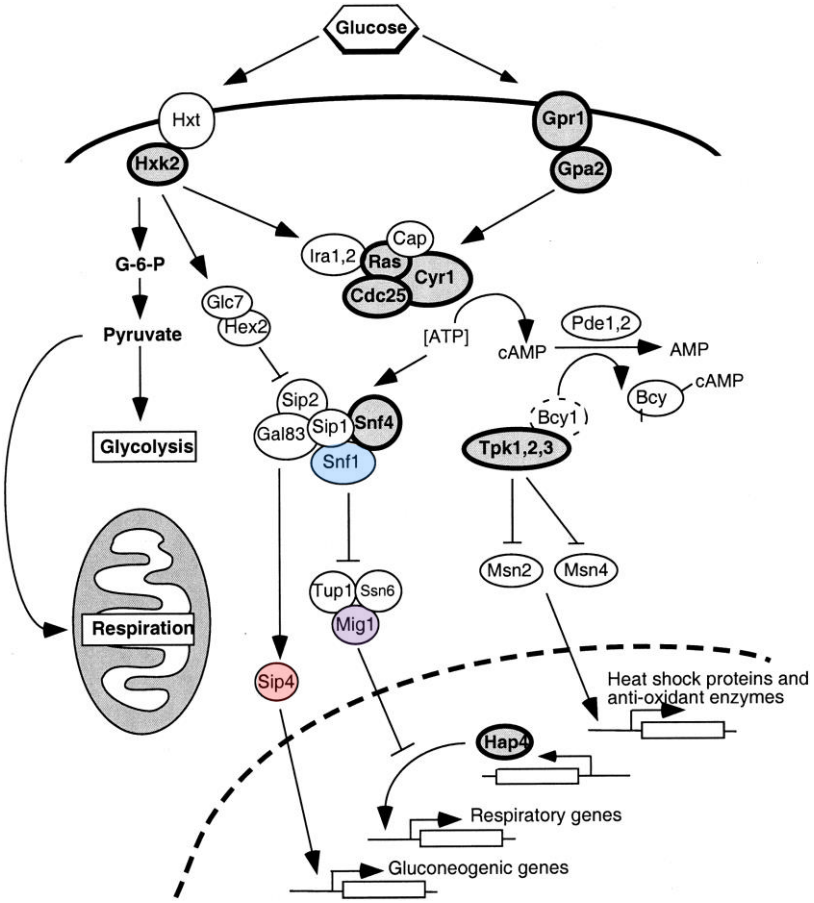
- Coordination of Research
- Experiments with the Robot Scientists ‘Eve’.
- Domain expertise - biological knowledge.

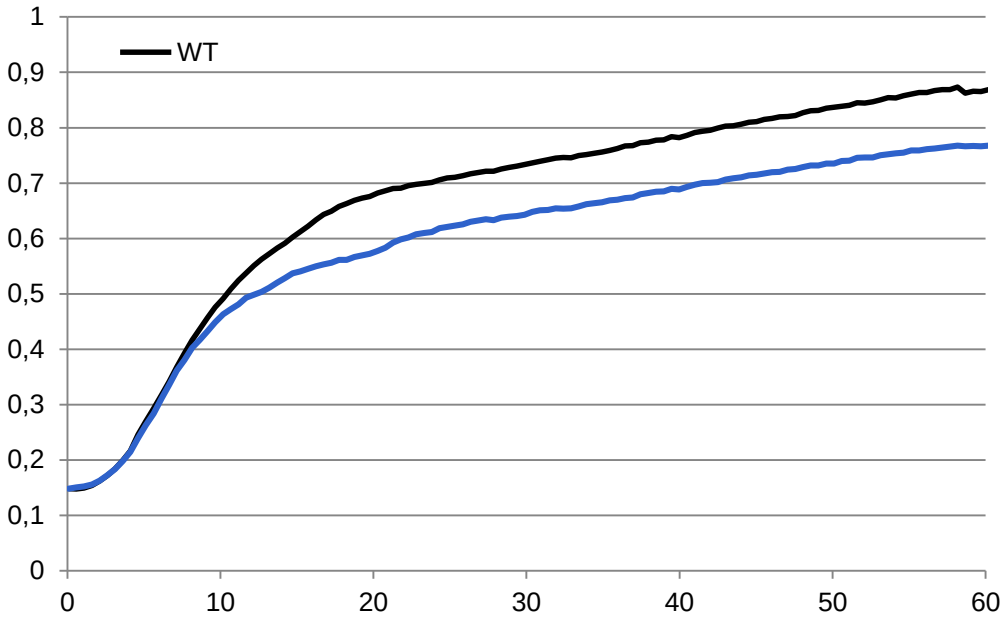
Key Outputs – Manchester

Automated Scientific Discovery

- Wet Lab –
  - Quantitative diauxic shift experiments have been successfully undertaken
- Experiment analysis software
  - Outputs experimental data in agreed format
  - Also outputs graphs and curve parameters
  - Can analyse multiple plates at once

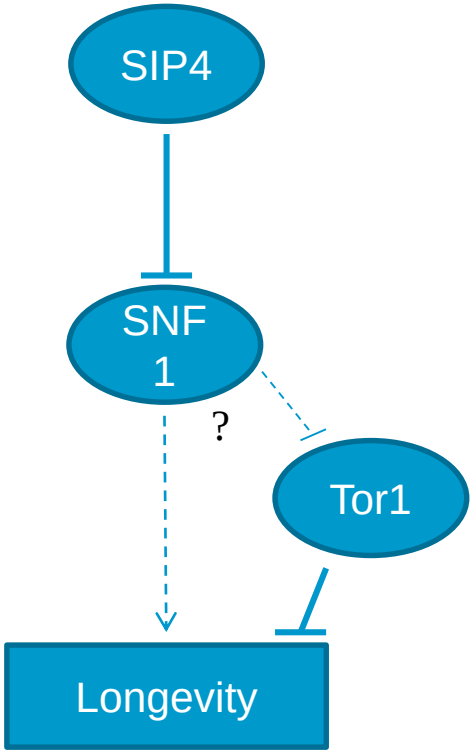
MIG2





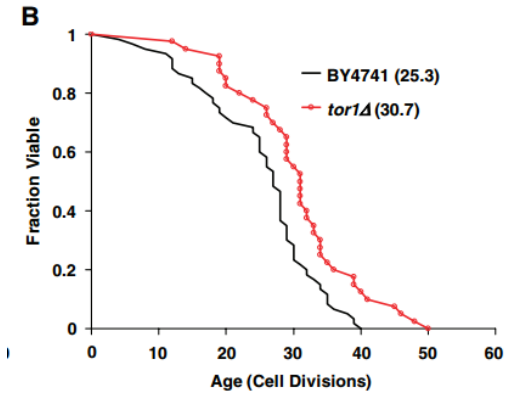
- Involved in response to low glucose conditions
- Strong similarities to human genes EGR1, EGR2, EGR2 and EGR4
- EGR2 is involved in PTEN-mediated apoptosis
- Human homolog WT1 mutated in Wilms' Tumour
- With MIG1, MIG2 is also regulated by SNF1

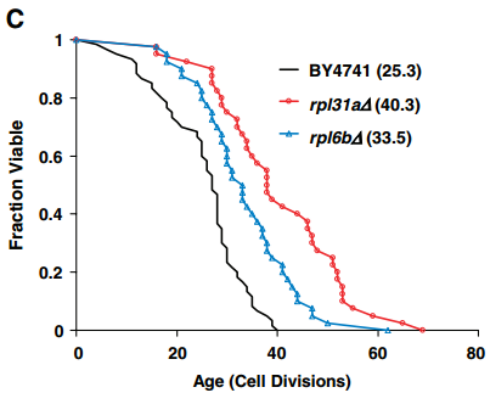
Bishop & Guarente, 2007



SIP4 & longevity

- SIP4 has been implicated in longevity in yeast, via Tor1
- Tor1 & downstream gene knockouts in BY4741 are able to survive more cell divisions than WT





Kaeberlein et al, 2005

Key Outputs - All

Experiment vs Simulation

- Integration of the components for learning, simulation, revision and experimental design into a global adaptive laboratory system is taking shape
- Interesting strains identified by simulators
- Mass Experimental Screening for interesting behaviour
- Work ongoing
  - Currently around 80 noteworthy strains covered

Key Outputs - summary

- A novel ontology for modelling uncertain knowledge – 90%
- An efficient communication mechanism between human and robot scientists – 50%
- New machine learning methods for the generation and efficient testing of complex scientific – 70%.
- Novel biomedical knowledge about cell biology relevant to cancer and ageing – 30%.

- AdaLab system – 70%

Next Steps

Next Steps

- **WP2:** Produce final versions of the knowledge representation models and the communication protocol.
- **WP3:** Finalise works on the integrated AdaLab framework and the application of Adalab for the investigation of the Warburg effect in cancer, and of calorie restriction in ageing using yeast as a model system.
- **WP4:** Improve the revision and experiment design strategies and to evaluate and validate the learning results obtained so far with real experiments
- **Overall:** to make all the Adalab components work together and to evaluate the whole Adalab workflow

chist-era

AdaLab

Potential Impact

Potential Impact

- Science is the greatest generator of economic wealth (through developments in technology).
- Science is the greatest driver of better health (through development in biomedical science).

- The AdaLab system's goal to be >20% more efficient at discovering scientific knowledge (within a limited scientific domain) than human scientists alone.

- Talking to Industry.

Dissemination

Dissemination

- Public Engagement
  - London Science Museum Event
  - Interview for Belgian television channel
- Dissemination to non-specialist audiences
  - Project Website and promotional literature
  - Newsletters
  - Public lectures and seminars (e.g. Pint of Science; Pendle Science café; Summer School of Science)
- Dissemination via specialist media
  - Conference presentations (23 given to date)
  - Journal publications (7 generated to date)
- International Workshop - planned

Sustainability/Valorisation

Sustainability

- Industrial links
  - European Laboratory Automation companies
  - European Pharmaceutical companies.
- Several components:
  - Machine learnt models of yeast behaviour and diauxic shift
  - Machine learnt models for selecting new experiments
  - AdaLab ontologies
  - Database of yeast growth data
  - Communication mechanisms
  - Software components
  - The integrated Adalab system

Valorisation

- 5 Postdocs training.
- 1 PhD student.
- 1 Masters student

- Social Science
  - 1 PhD student sociology
  - 1 PhD student anthropology

Thanks