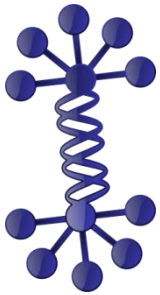
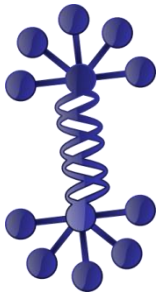


Ondex – Data integration and visualisation

The Ondex project

contact@ondex.org

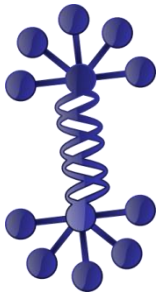
BBSRC bioscience for the future



Outline

- Data integration in Systems Biology
- The Ondex approach
- Application cases

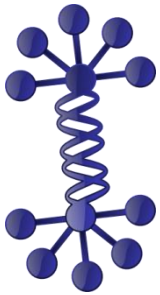
BBSRC
bioscience for the future



Outline

- Data integration in Systems Biology
- The Ondex approach
- Application cases

BBSRC
bioscience for the future



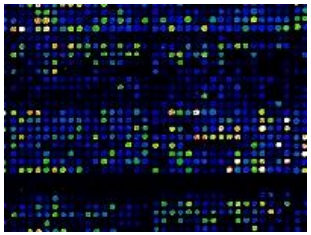
BBSRC defines Systems Biology

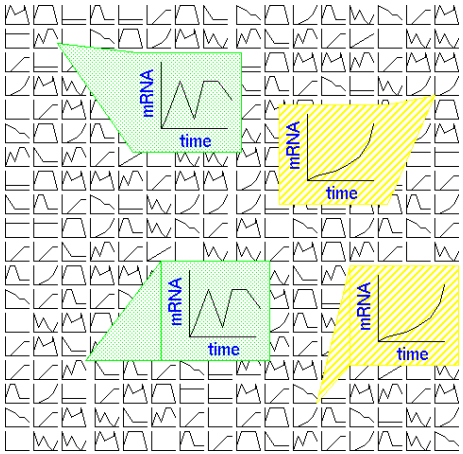
"Systems biology depends on high-powered computation to construct predictive models.

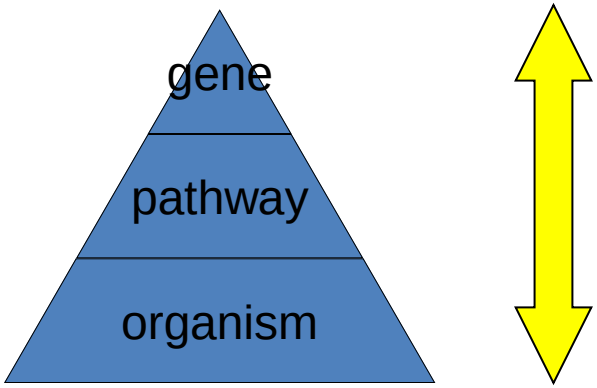
Typically these draw upon the large amounts of quantitative data generated by high-throughput techniques such as genomics (DNA content of cells), transcriptomics (the messenger RNA produced from active genes), proteomics (the proteins produced from the messenger RNA) and metabolomics (small compounds in cells).

The models relate these data to research on metabolic and other functions in cells and tissues, and to the physiology and behaviour of whole organisms."

Opportunities and challenges for modelling



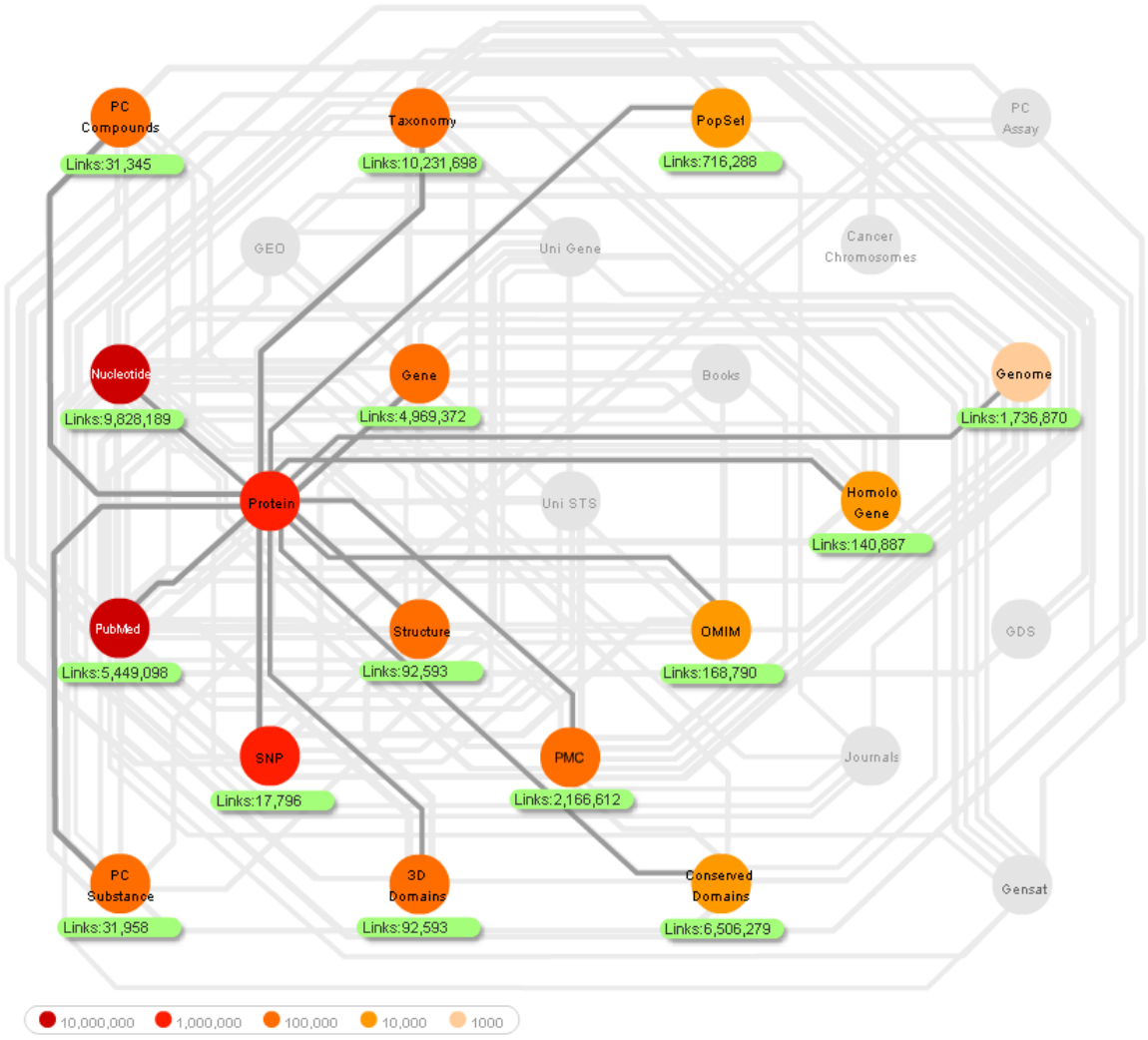




- Large amounts of multi-'omics data
  - Genome, microarray, metabolomics, PPI …
- The biological systems span multiple levels of biological organisation
- Non-trivial to integrate the data

BBSRC bioscience for the future

Syntactic integration challenge



http://www.ncbi.nlm.nih.gov/Database

Over **1000 databases** freely available to public

Over **60 million sequences** in GenBank

Over **870 complete genomes** and many ongoing projects

Over **17 million citations** in PubMed

PubMed growth by 600,000 publications each year

Integration of Life Science data sources is essential for Systems Biology research

BBSRC bioscience for the future

Semantic Integration challenge

- Same concept different names
  - synonyms
  - ontologies
- Same name different concepts
  - homographs

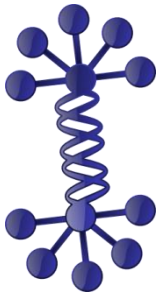
Plant AmiGO Ontology DATABASE: PO_0205

inflorescence

Accession: PO:0009049
Aspect: plant structure
Synonyms:
cob (sensu sorghum)
corymb
cyme
dichasium
drepanium
helicoid cyme
monochasium
panicle
raceme
rhipidium
scorpioid cyme
spike (sensu Triticeae)
umbel
verticillaster
Definition:
That part of the axial system of plants above the uppermost foliage leaf/pair of foliage leaves that bears flowers.
Comment:
Some plants have only solitary flowers, e.g. Magnolia.

Ear

BBSRC bioscience for the future

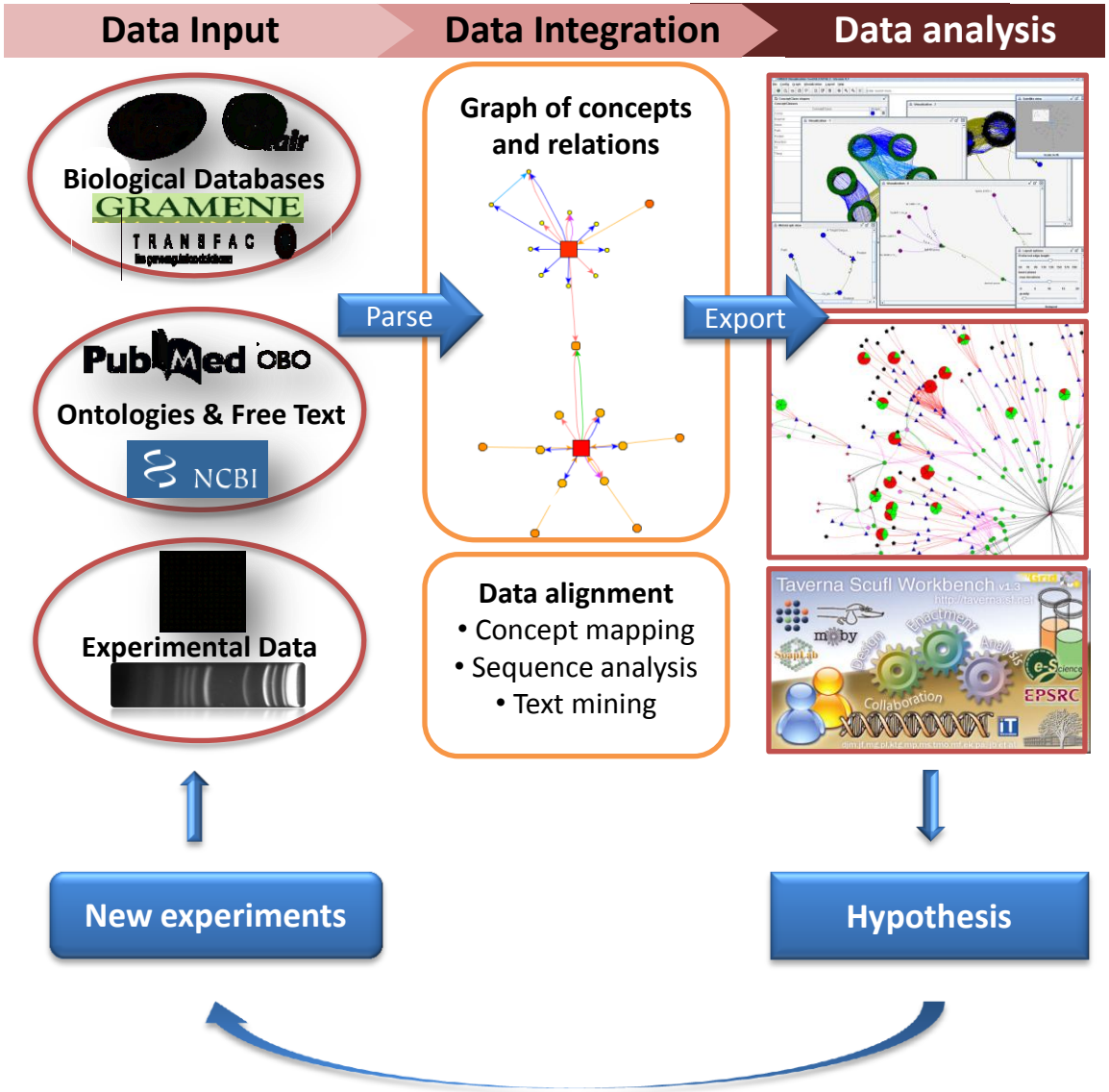


Outline

- Data integration in Systems Biology
- The Ondex approach
- Application cases

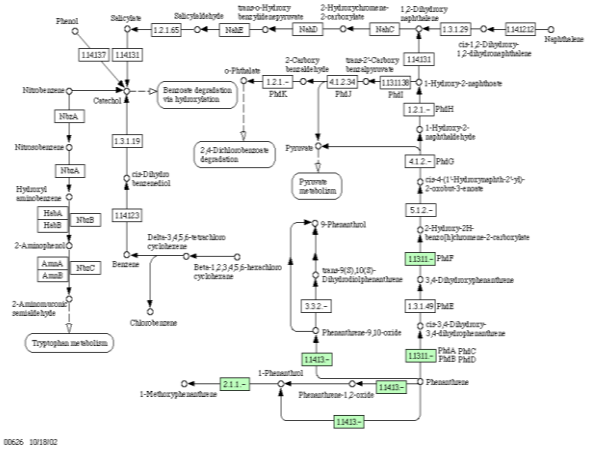
BBSRC
bioscience for the future

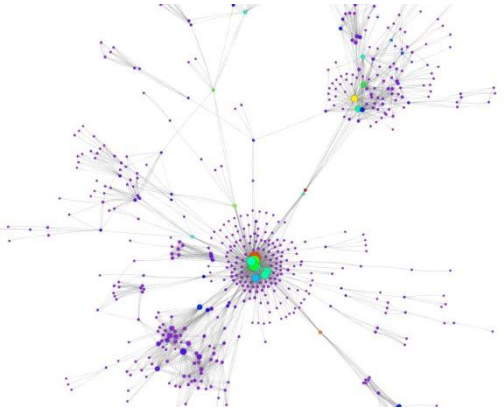
Data integration and visualisation in Ondex

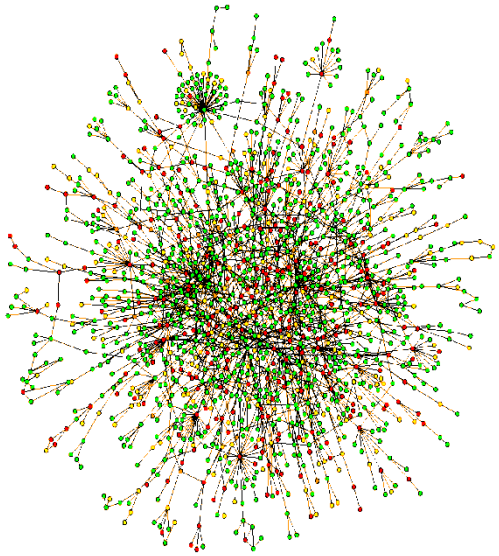


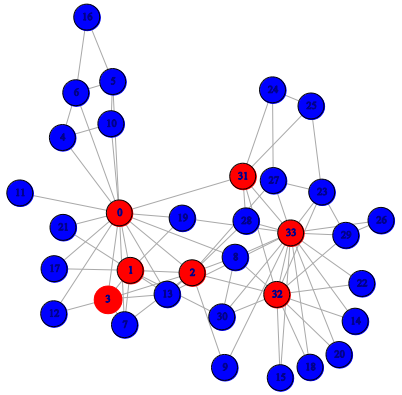
BBSRC bioscience for the future

Everything is a network







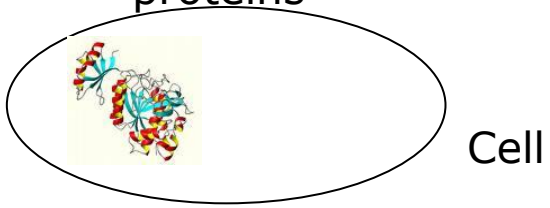


BBSRC
bioscience for the future

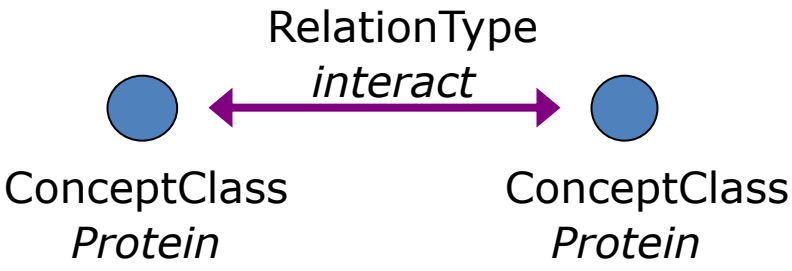
Semantic networks in Ondex

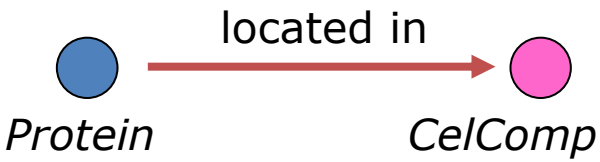
Protein – Protein interaction network (PPI)

Cellular location of proteins



Ondex: Network of Concepts and Relations



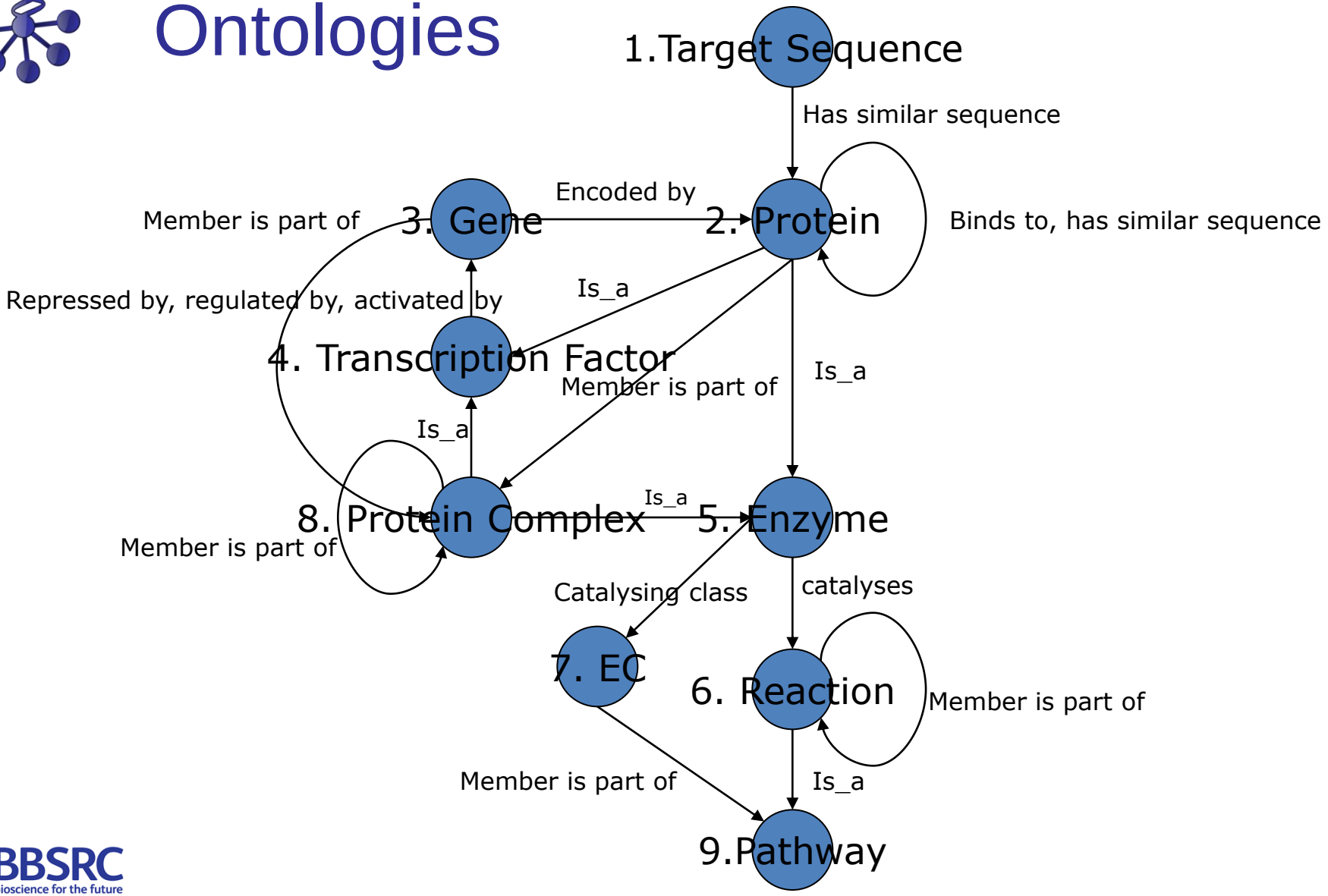


Properties: compound name, protein sequence, protein structure, cellular component, KM-value, PH optimum …

Ontology of Concept Classes, Relation Types and additional Properties

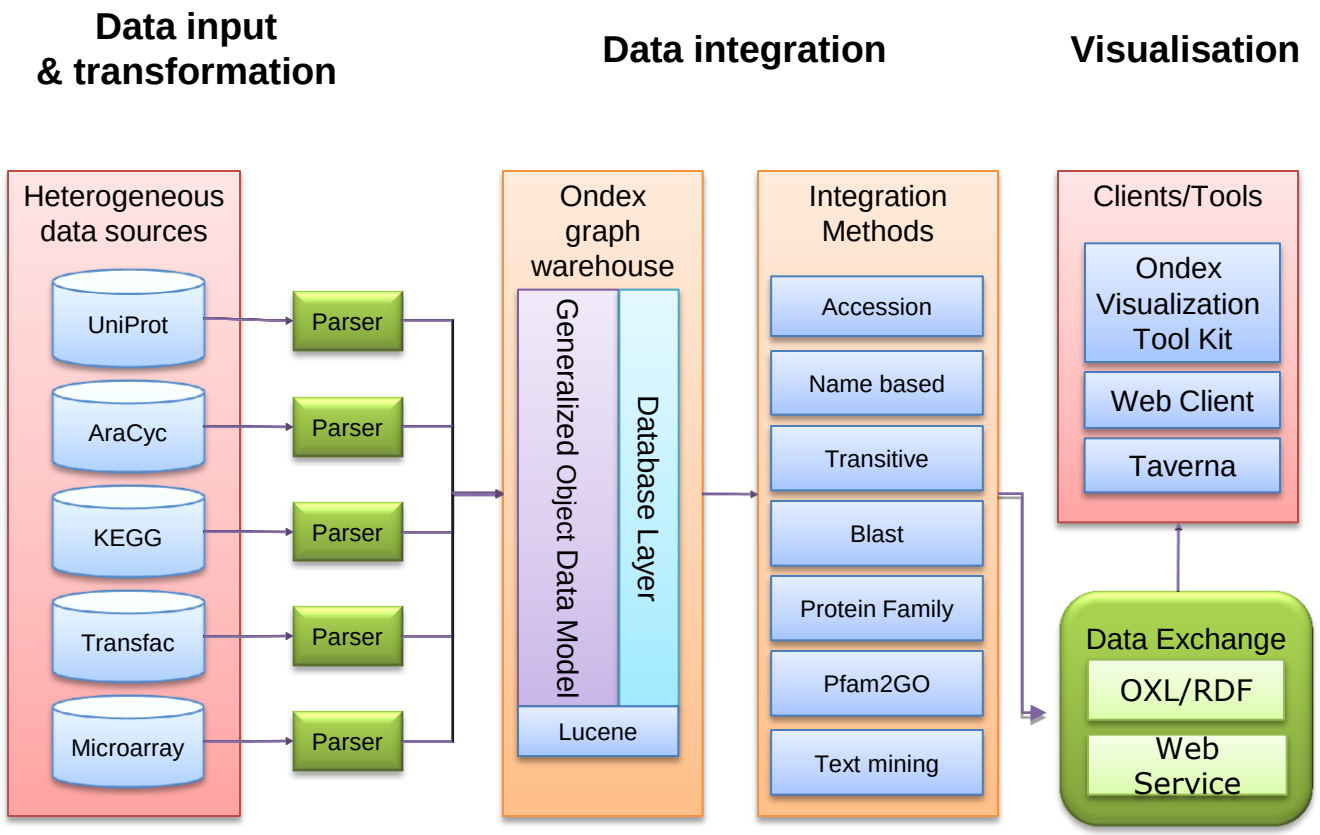
BBSRC bioscience for the future

Metagraph – based on upper level Ontologies

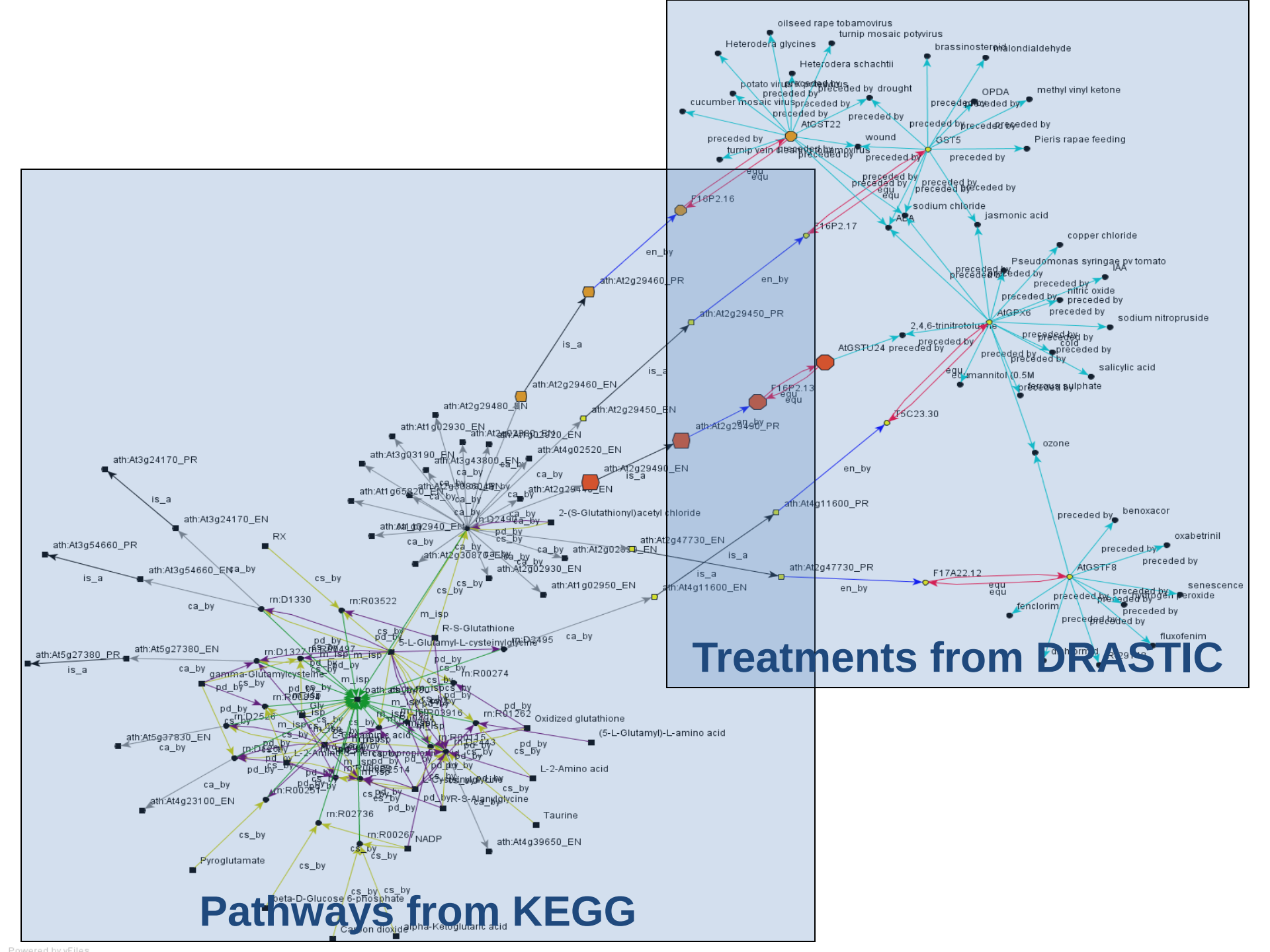


BBSRC bioscience for the future

Ondex Data Integration Scheme

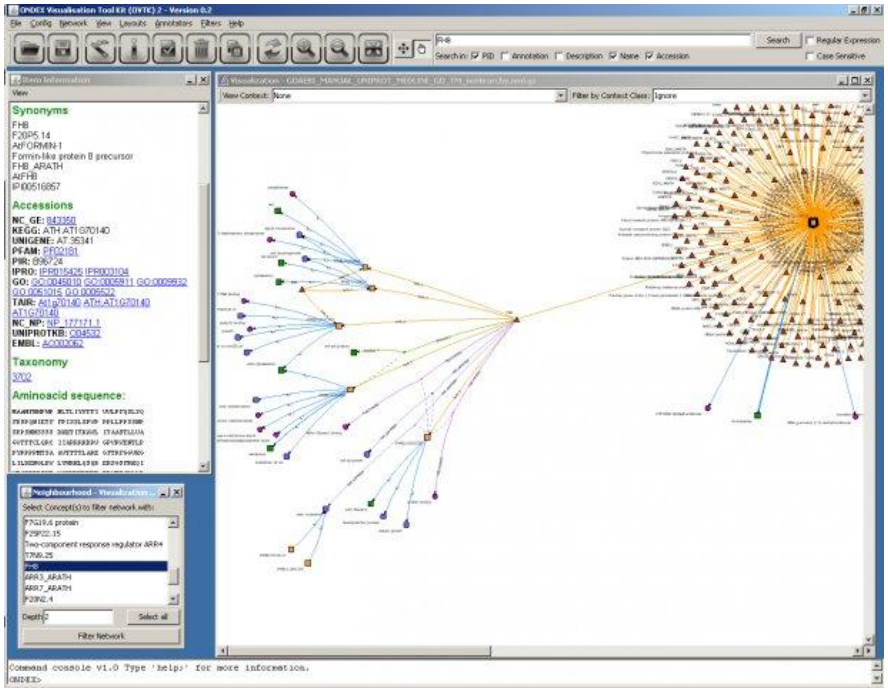


BBSRC bioscience for the future


oilseed rape tobamovirus
turnip mosaic potyvirus
Heterodera glycines
brassinosteroid
malondialdehyde
Heterodera schachtii
potato virus
preceded by drought
OPDA
methyl vinyl ketone
cucumber mosaic virus
preceded by
AtGST22
GST5
wound
Pieris rapae feeding
turnip vein clearing virus
F16P2.16
F16P2.17
sodium chloride
jasmonic acid
ABA
copper chloride
Pseudomonas syringae pv tomato
IAA
nitric oxide
AtGPX6
sodium nitroprusside
2,4,6-trinitrotoluene
AtGSTU24
cold
salicylic acid
mannitol (0.5M
ferrous sulphate
T5C23.30
ozone
ath:At2g29460_PR
ath:At2g29450_PR
ath:At2g29460_EN
ath:At2g29480_EN
ath:At2g29450_EN
F16P2.13
ath:At2g29490_PR
ath:At2g29490_EN
ath:At4g11600_PR
ath:At2g47730_PR
ath:At2g47730_EN
ath:At4g11600_EN
F17A22.12
AtGSTF8
benoxacor
oxabetrinil
senescence
hydrogen peroxide
fenclorim
fluxofenim
ath:At3g24170_PR
ath:At3g24170_EN
ath:At3g54660_PR
ath:At3g54660_EN
RX
rn:D1330
rn:R03522
2-(S-Glutathionyl)acetyl chloride
R-S-Glutathione
5-L-Glutamyl-L-cysteinylglycine
gamma-Glutamylcysteine
ath:At5g27380_PR
ath:At5g27380_EN
rn:R00274
rn:R01262
Oxidized glutathione
(5-L-Glutamyl)-L-amino acid
ath:At5g37830_EN
L-2-Amino acid
ath:At4g23100_EN
rn:R02736
Taurine
rn:R00267
NADP
ath:At4g39650_EN
Pyroglutamate
beta-D-Glucose 6-phosphate
Carbon dioxide
alpha-Ketoglutaric acid
Treatments from DRASTIC
Pathways from KEGG

From data integration to data visualisation

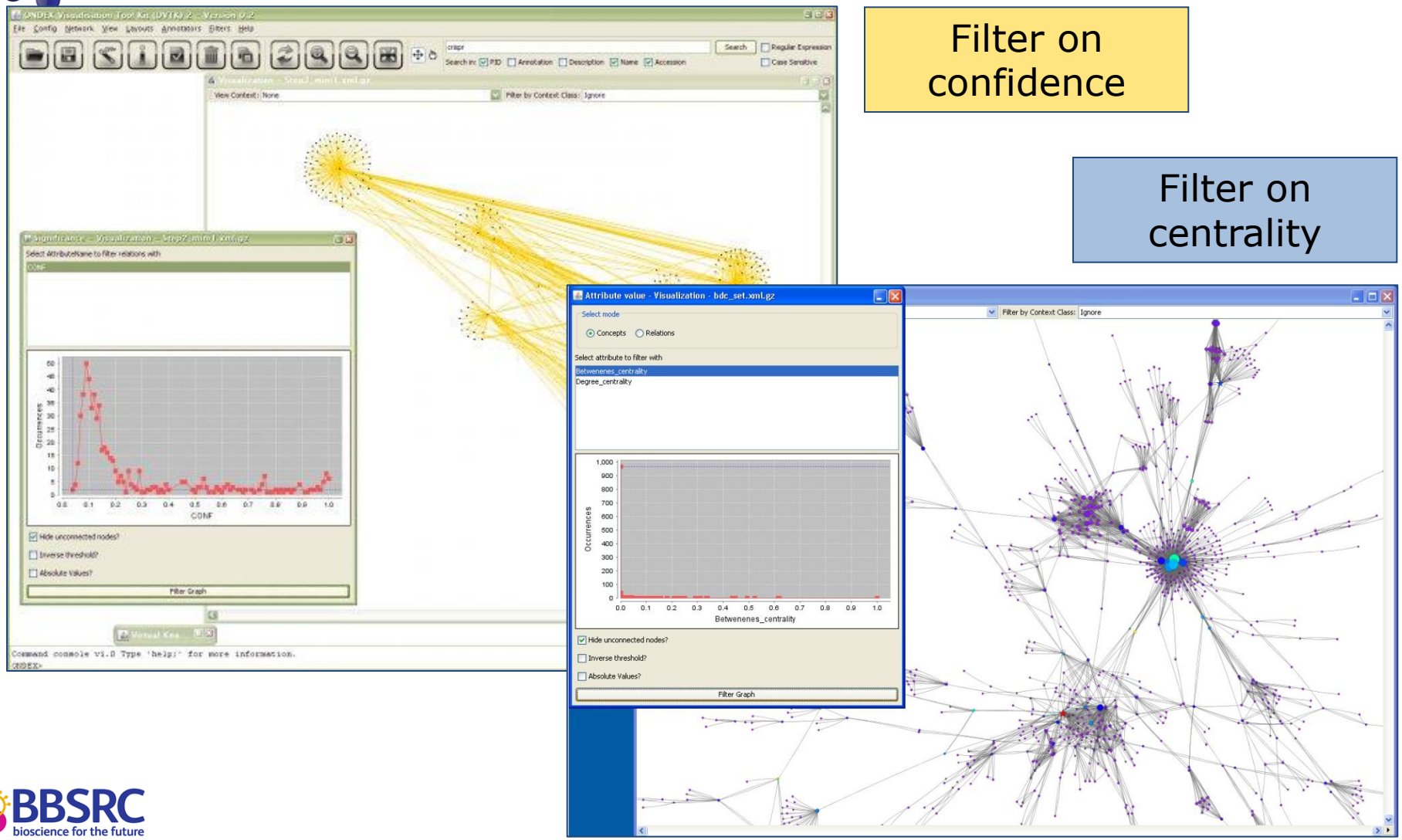
- Addressed issue of interrelatedness of databases
  - Syntactic integration
  - Semantic integration



- Complexity of interactions
  - PPI, co-expression, co-citation, …
- Bring together data, exploit graph structure
  - Candidate gene prioritisation and pathway discovery

BBSRC bioscience for the future

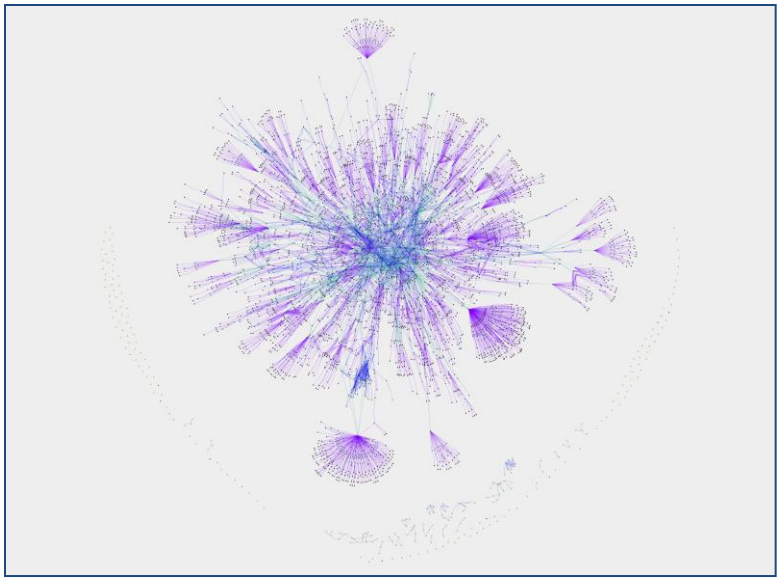
Filter on network properties



BBSRC bioscience for the future

Exploiting network representations

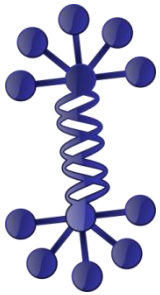
- The network is more than a visual dump of the database
- Can it be used predictively?
- Can we combine different types of information (evidence) with a network-based approach?



What are the important regions in the network?

Can we use graph theoretical techniques to identify nodes / regions of the network?

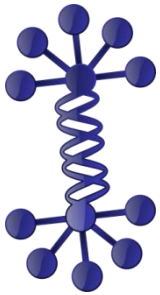
BBSRC bioscience for the future



Technical Philosophy

- Domain independent approach to semantic data integration
  - Focussed on biological problems
- Open source
- Open standards where available
  - SBML, BioPAX, PSI-MI, OBO
- Open interfaces for data exchange
  - Web services
  - Data export

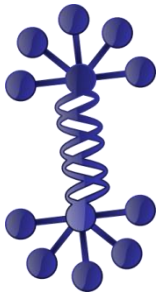
BBSRC bioscience for the future



Outline

- Data integration in Systems Biology
- The Ondex approach
- Tutorial

BBSRC
bioscience for the future



Acknowledgements

Rothamsted members:

- Catherine Canevet
- Keywan Hassani-Pak
- Stephen Hanley
- Matthew Hindle
- Angela Karp
- Shao Chih Kuo
- Artem Lysenko
- Chris Rawlings
- Mansoor Saqi
- Andrea Splendiani
- Jan Taubert

Former members:

- Jacob Köhler
- Rainer Winnenberg

Manchester members:

- Sophia Ananiadou
- Paul Dobson
- Paul Fisher
- Carole Goble
- Gina Levow
- Pedro Mendes
- Raheel Nawaz
- Georgina Moulton
- Robert Stevens
- David Withers
- Katy Wolstencroft

Edinburgh members:

- Igor Goryanin
- Andrew Millar
- Luna De Ferrari

Newcastle members:

- Simon Cockell
- James Dewar
- Eva Holstein
- Katherine James
- Philip Lord
- David Lydall
- Matthew Pocock
- Jochen Weile
- Darren Wilkinson
- Anil Wipat

Biological collaborators:

- Kim Hammond-Kosack
- Martin Urban
- Dimah Habash
- David Wild
- Katherine Denby
- Roxane Legaie

BBSRC
bioscience for the future