

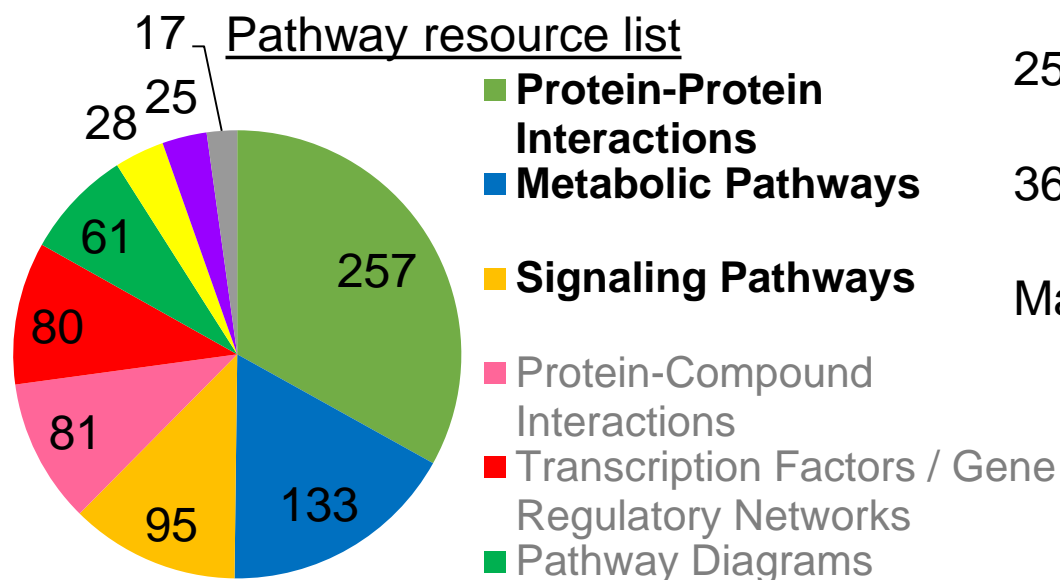
# Shared Resources, Shared Costs Leveraging Biocuration Resources

Sandra Orchard  
orchard@ebi.ac.uk

# Why resource sharing is essential

- Manual curation is essential to extracting biological meaning from the wealth of data available to us
- Manual curation is expensive and requires expensive custom software - databases and editorial tools
- Redundant curation, redundant software development and data loss when resources are lost are therefore a luxury the community cannot afford
- We need to collaborate, not compete, to manage this better

# Molecular interaction resources



257 PPI resources in 2013

36 completely unavailable

Many more not updated for >2 years

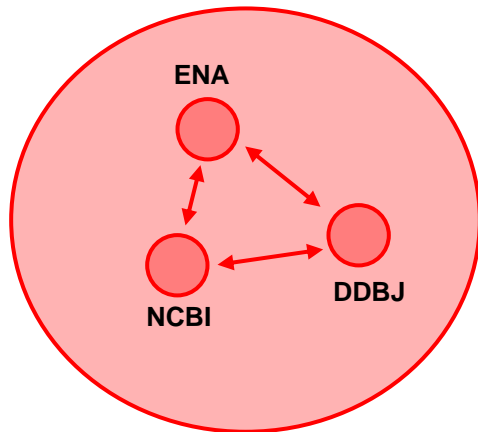
**547 resources – but how many are actively maintained and added to?**

Pathguide» the pathway resource list <http://www.pathguide.org/>

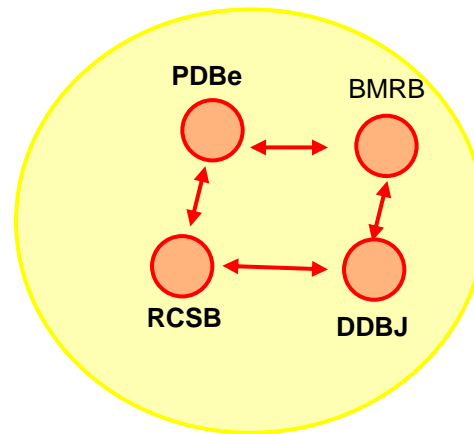
# Collaboration is not a new idea

- More data coverage
- Less redundancy
- Less inconsistency
- Better data management

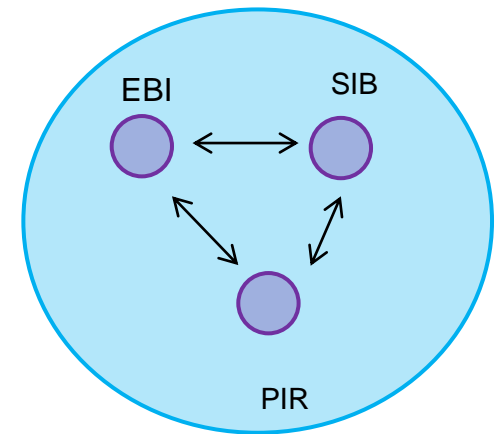
Access, exchange, sharing, portability, interoperability, annotation, comparison, verification, representation, integration, reusability.



Nucleotide sequences  
**INSDC**



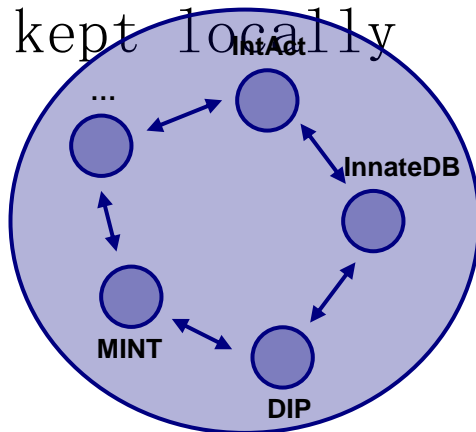
Protein Structures  
**wwPDB**



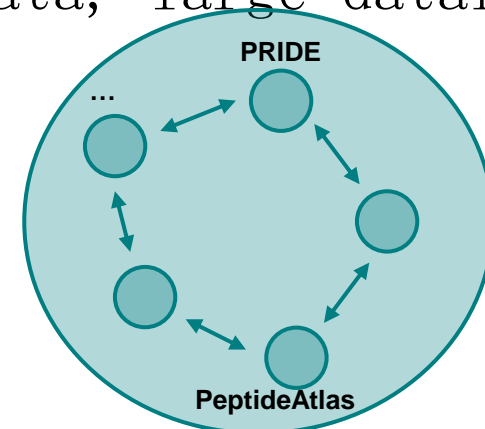
Protein Sequences  
**www.uniprot.org**

# One mechanism does not fit all

- INSDC and wwPDB - file exchange
- UniProt - single database
- IMEx - use web services to populate a single website
- ProteomeXchange - share metadata, large datafiles kept locally



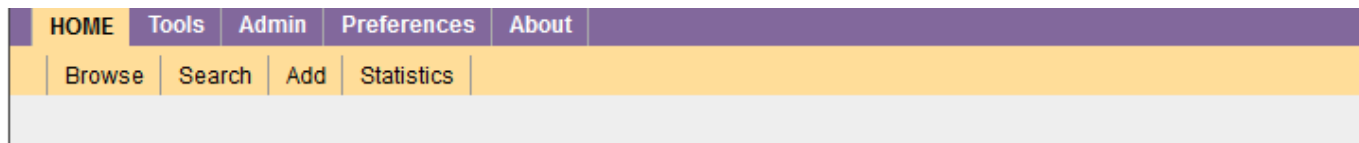
Molecular interactions  
**IMEx**



Protein identifications  
**ProteomeXchange**



# Resource sharing - minimises curation redundancy



## Imex Central News

Recent News:

2014-09-20

ImexCentral v 1.03: Improved...

2014-05-11

ImexCentral v 1.02: Record...

2014-01-26

ImexCentral v 1.01: Support for...

See more [Imex Central News](#)

## Welcome to ImexCentral

IMEx Central is designed to coordinate curation effort and the exchange of completed records on molecular interaction data providers that agreed to conform to common curation and data exchange formats. We hope that this way, by users with a large, uniform set of biological interactions. Such dataset set will be easier to analyse than a most incompatible ways.

The databases actively participating in the coordinated data curation currently include:

- [DIP](#)
- [InnateDB](#)
- [IntAct](#)
- [MatrixDB](#)
- [MINT](#)
- [Molecular Connections](#)
- [Uniprot Consortium](#)
- [UCL-BHF](#)

Decision about publication already in IMEx central



This publication is already curated by a IMEx partner. What do you want to do?

Create the publication anyway

Do not create the publication

# Data sharing prevents data loss

- PIR had lost funding when it joined UniProt - database contained sequences which were not present in Swiss-Prot or TrEMBL
- Microbial Protein Interaction Database (MPIDb) - PI moved into a new field, data imported and maintained by IntAct as part of the IMEx Consortium



# Important considerations

- Acknowledging contribution - collaborations only work if the recognition is shared.
- Gathering metrics - funding agencies require these for each grant/institution so this capability needs to be built into any system.
- Tool developers need to gather ideas from everyone - don't build something that only suits the host group's working practices
- Be prepared to compromise - in the long run you gain more than you lose
- Competitors are much more useful when they become collaborators



MINT



Thank you!

