

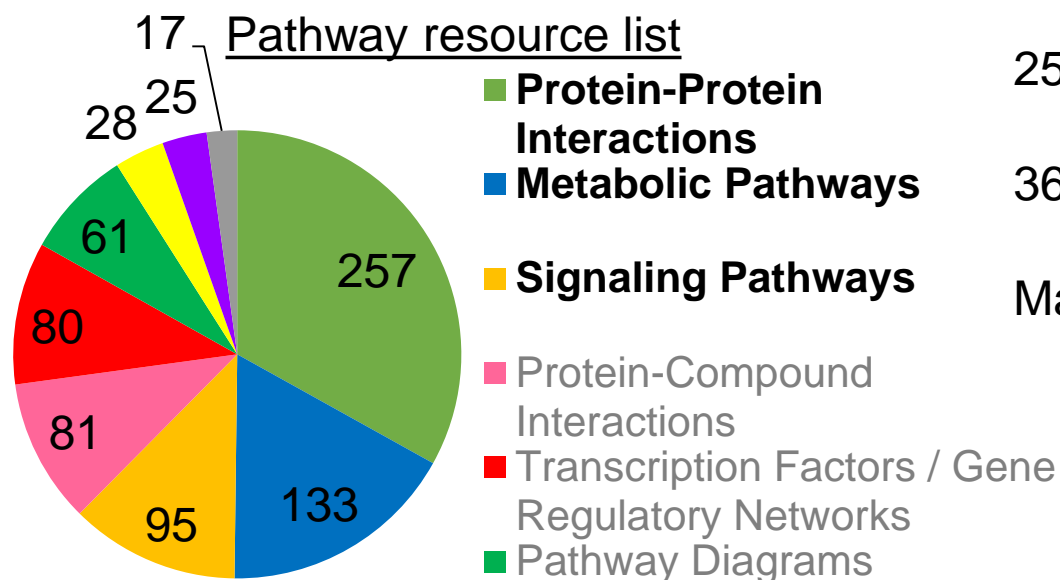
Shared Resources, Shared Costs Leveraging Biocuration Resources

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

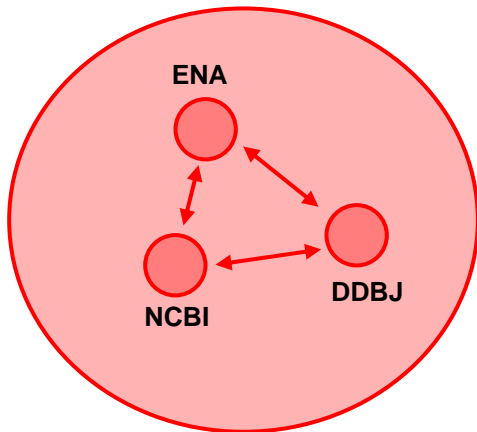
547 resources – but how many are actively maintained and added to?

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

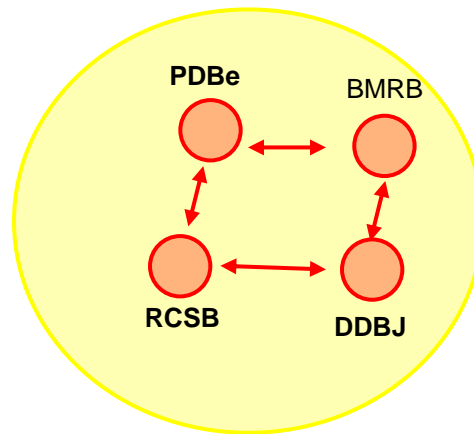
Collaboration is not a new idea

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

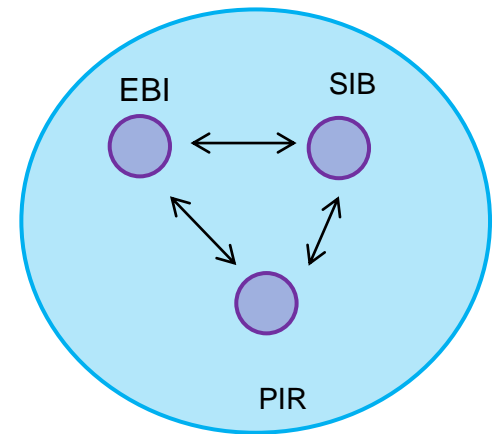
Access, exchange, sharing, portability, interoperability, annotation, comparison, verification, representation, integration, reusability.



Nucleotide sequences
INSDC



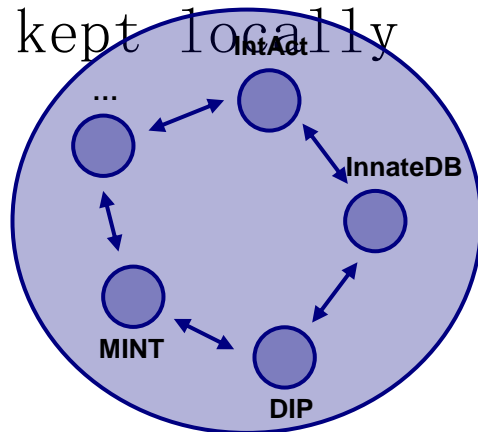
Protein Structures
wwPDB



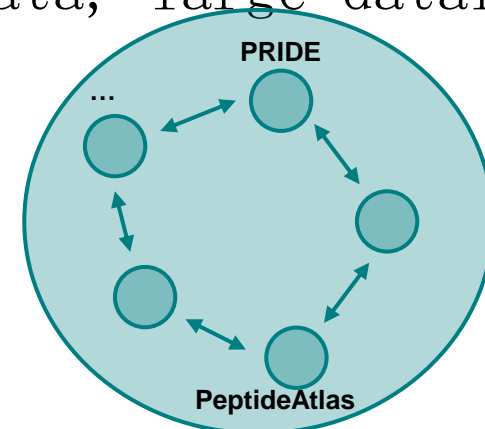
Protein Sequences
www.uniprot.org

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 software development

- IntAct - one database, one editor, 13 institutes

Publications owned by you (26) Publications reviewed by you (23) All the publications (141)												
Pub. ID	First author	Publication year	Title				# Exps.	# Inter.	Owner	Reviewer	Status	
25525879	Mathias RA. et al.	2014	Sirtuin 4 is a lipamidase regulating pyruvate dehydrogenase complex activity.				7	9	orchard	mgt	CURATION_IN_PROGRESS	
24630722	Lu A. et al.	2014	Unified polymerization mechanism for the assembly of ASC-dependent inflammasomes.				9	21	orchard	pporras	CURATION_IN_PROGRESS	
25594178	Tan X. et al.	2015	A kinase-independent role for EGF receptor in autophagy initiation.				7	33	orchard		CURATION_IN_PROGRESS	
25525882	Cho SH. et al.	2014	Detecting envelope stress by monitoring β -barrel assembly.				5	9	orchard		READY_FOR_CHECKING	
25467444	Freund A. et al.	2014	Proteostatic control of telomerase function through TRiC-mediated folding of TCAB1.				6	7	orchard	pporras	READY_FOR_CHECKING	
25416944	Joachimiak L.A. et al.	2014	The structural basis of substrate recognition by the eukaryotic chaperonin TRiC/CCT.				3	6	orchard	mgt	CURATION_IN_PROGRESS	
unassigned1260			FIH regulates cellular metabolism through hydroxylation of the deubiquitinase OTUB1.				0	0	orchard		CURATION_IN_PROGRESS	
unassigned1259			VRK1 chromatin kinase phosphorylates H2AX and is required for foci formation induced by DNA damage				2	2	orchard		CURATION IN PROGRESS	
25417162	Sarma K. et al.	2014	ATRX directs binding of PRC2	keller	Guillaume	Keller	●	UniProt		CURATOR		
				rhieta	Reija	Hieta	●	UniProt		CURATOR		
				laksrp	Lakshmi	Pillai	●	HPIDb		CURATOR		
				ammari	Mais	Ammari	●	HPIDb		CURATOR		
unassigned1242			GADD34 Facilitates Oxidative Disease	lperetto	Livia	Peretto	●	MINT		REVIEWER, CURATOR		
				cesareni	Gianni	Cesareni	●	MINT		CURATOR		
			molecular sieving	2040637		physical association	EBI-6421327		IntAct	COMPLEX_CURATOR, REVIEWER, CURATOR		
			molecular sieving	9545254		physical association	EBI-6453429		IntAct	Curator Database		
ftsZ	P0A9A6 EBI-370963	ftsZ	P0A9A6 EBI-370963	Escherichia coli (strain K12)	Escherichia coli (strain K12)	adenylate cyclase complementation	18394147 imex : IM-18198		physical association	EBI-6403140 MPIDB-INT-947 imex : IM-18198-8	mpidb	ADMIN CURATOR
			electron microscopy	21216995		physical association	EBI-6413618 MPIDB-INT-1532 imex : IM-18383-2		mpidb	CURATOR ADMIN, CURATOR		
			electron microscopy	21216997		direct interaction	EBI-6452373 imex : IM-18384-6		mpidb	ADMIN, COMPLEX_CURATOR, REVIEWER, COMPLEX_REVIEWER, CURATOR		
			light scattering	21216997		direct interaction	EBI-6452363 imex : IM-18384-5		mpidb			
			cosedimentation	22515815		self interaction	EBI-6384758 imex : IM-17992-5		UniProt			
			cosedimentation	19415799		physical association	EBI-6407788 MPIDB-INT-671 imex : IM-18288-4		mpidb			
			light scattering	22064072		direct interaction	EBI-7518474 MINT-2204111		MINT			

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



MB:Info
A modular approach to cellular functions



SGD *Saccharomyces*
GENOME DATABASE



Thank you!

