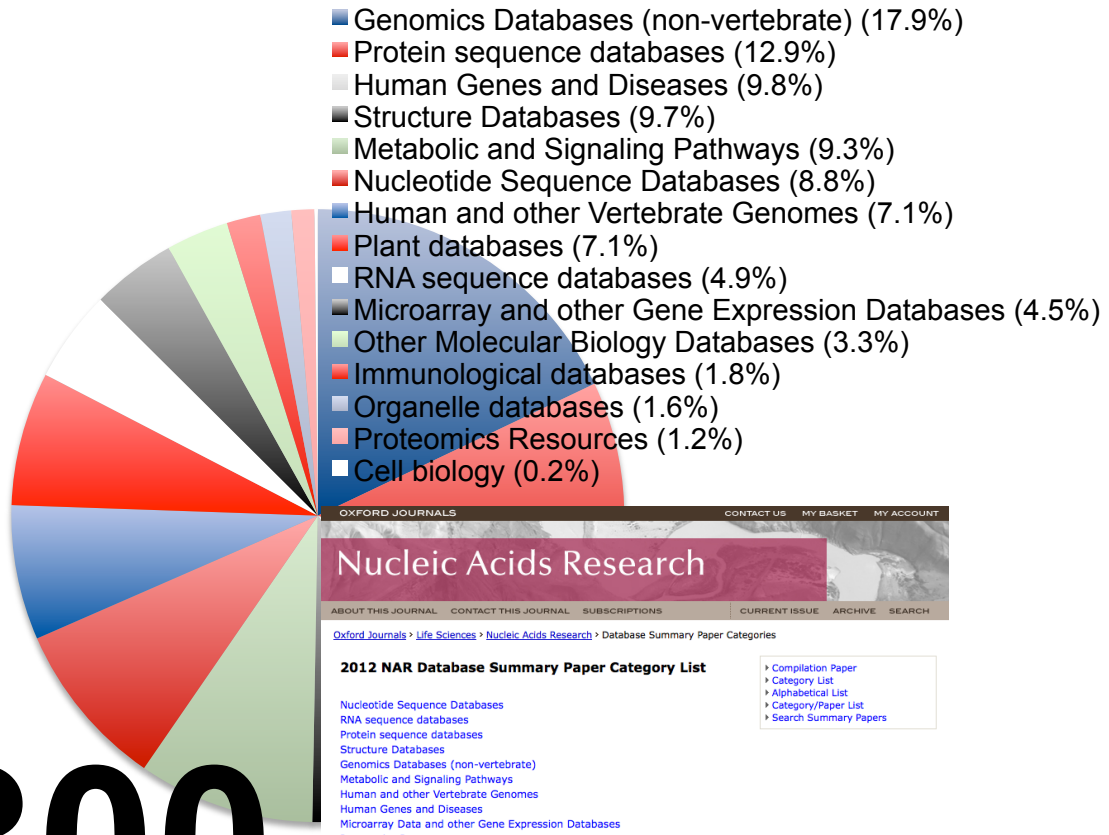
Minimum metadata to interlink services in life sciences

Manuel Corpas, Project Leader at TGAC and ELIXIR UK deputy technical coordinator

Life sciences data sources

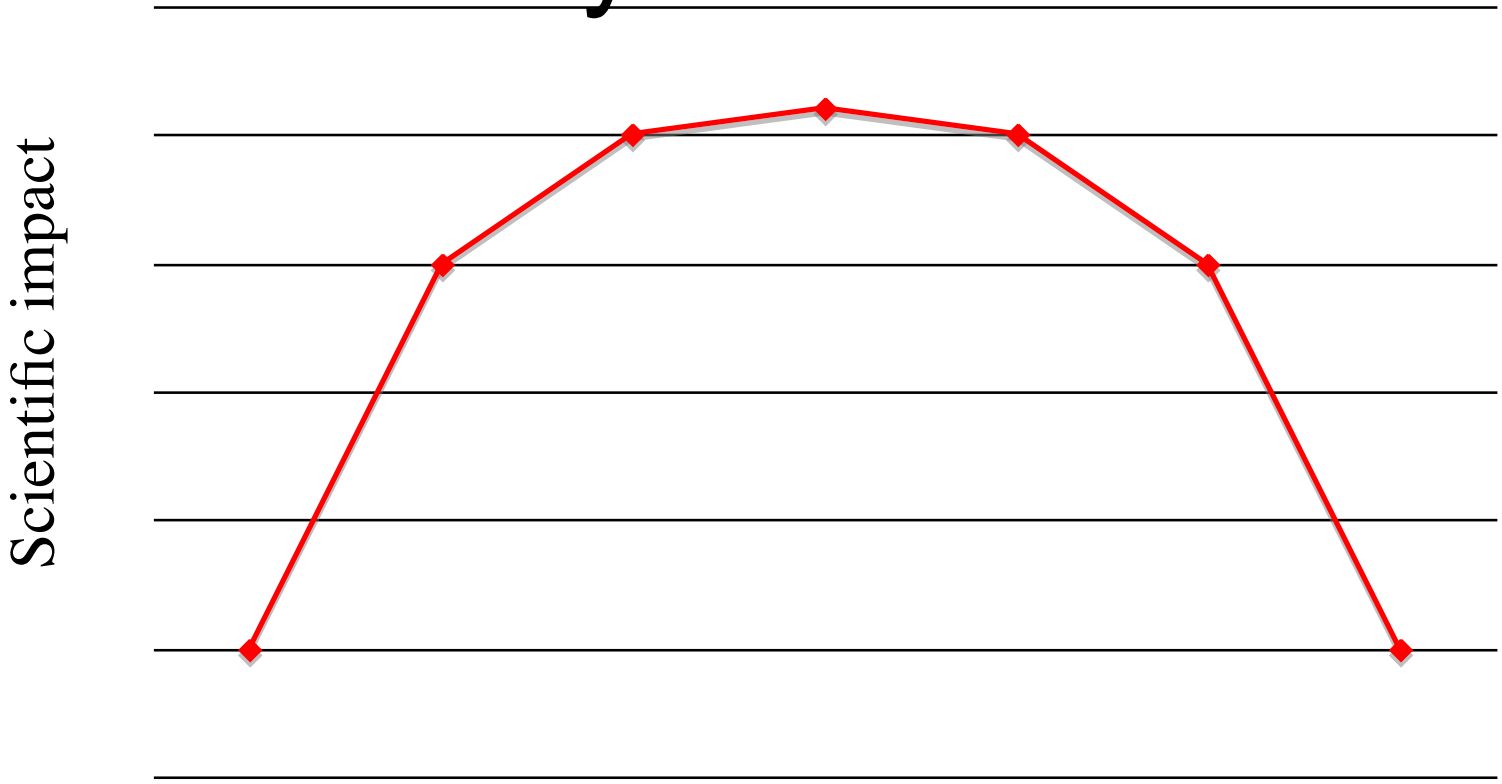
- Diverse
- Disperse
- Massive



~1800
molecular biology data resources

Nucleic Acids Research annual Database Issue
and the NAR online Molecular Biology Database Collection in 2012.
MY Galperin, GR Cochrane – Nucleic Acids Research, 2011

Utility of databases

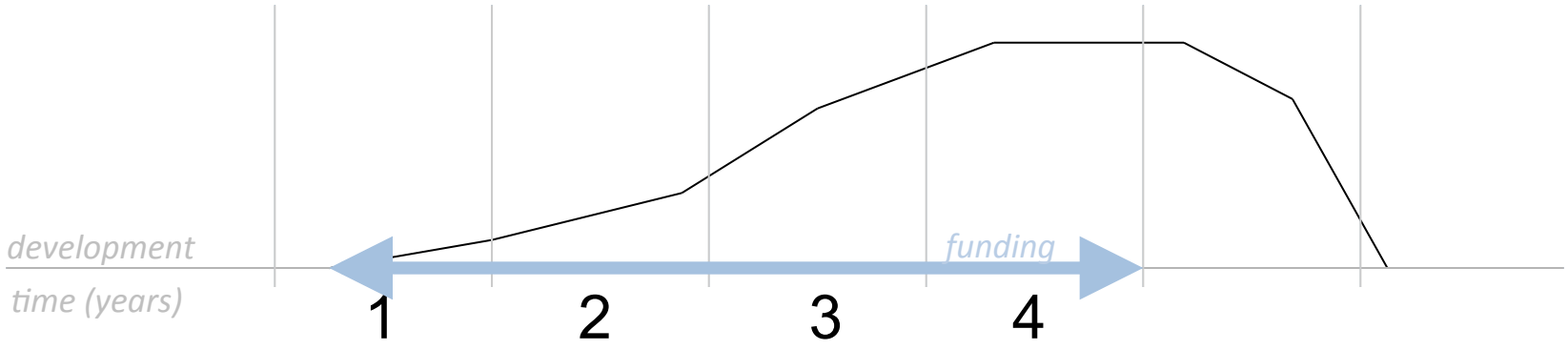


Why so many sources?

- Complexity of data
- Many communities
- Funding encourages new development

Are they sustainable?

- Just 20% has a sustained future*



** Merali Z. et all. Databases in peril. Nature 2005.*

This is not a unique problem for the Life Sciences

amazon.co.uk

Manuel's Amazon | Today's Deals | Gift Cards | Sell | Help

Shop by Department | Search | All | Go | Hello, Manuel Your Account

- Amazon Instant Video
- MP3s & Cloud Player
  20 million songs, play anywhere
- Amazon Cloud Drive
  5 GB of free storage
- Kindle
- Appstore for Android
  Get a premium app for free everyday
- Digital Games & Software
- Audible Audiobooks
  100,000 audiobook downloads

- Books
- Movies, TV, Music, Games
- Electronics & Computers
- Home, Garden, Pets & DIY
- Toys, Children & Baby
- Clothes, Shoes & Jewellery
- Sports & Outdoors
- Beauty, Health & Grocery
- Car & Motorbike

Full Shop Directory

Amazon Mobile Apps | Instant Video | Kindle | Cloud Drive | Appstore for Android | Audible Audiobooks

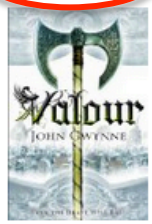
All-New
kindle fire HD
The perfect family tablet
from £119 › Shop now

kindle
Now save £10
Only £59 ~~£69~~ › Shop now

Amazon Family | Trade-In | From ice cream to electronics | Amazon Prime | Subscribe & Save | Clothing Store

Family business enables family dream
Ice cream seller changes direction and hits million-pound turnover
› Discover Julius Oliveti's story
One of thousands of small businesses thriving because of Amazon customers

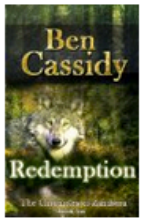
New for You



Valour (Faithful & the Fallen 2)
› John Gwynne
Kindle Edition
★★★★½ (35)
£6.59
Fix this recommendation

City of Light
› Will Wight
Kindle Edition
★★★★½ (3)
£1.85
Fix this recommendation

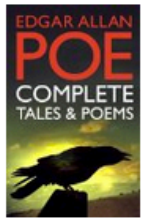
Master Mage (Reawakening Saga)
› D.W. Jackson
Kindle Edition
£2.98
Fix this recommendation



Redemption
› Ben Cassidy
Kindle Edition
★★★★☆ (2)
£1.85
Fix this recommendation

Requiem's Song (Dawn of Dragons, Book 1)
› Daniel Arenson
Kindle Edition
★★★★★ (1)
£1.99
Fix this recommendation

The Academy: Book 2
› Chad Leito
Kindle Edition
★★★★½ (2)
£2.47
Fix this recommendation



Edgar Allan Poe: Complete Tales and...
Edgar Allan Poe, Maplewood Books
Kindle Edition
★★★★☆ (36)
£0.77

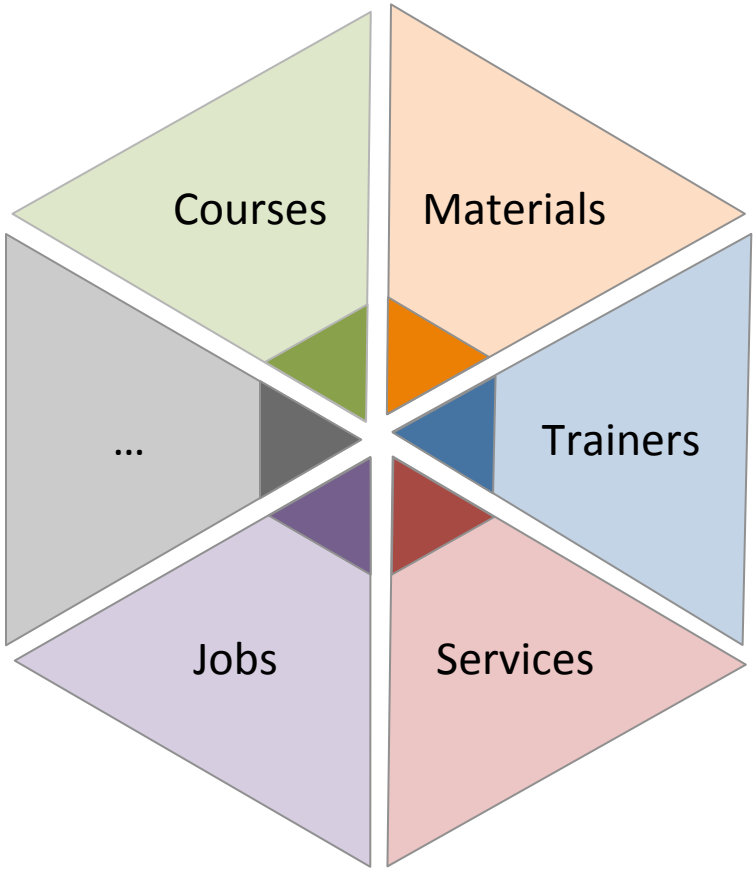
How could this be achieved?

- If organizations could agree on adopting a **minimum set of metadata** to annotate services and information we could **interconnect services**, **provide context information** and help users with **recommendations** the same way it is done in ecommerce websites

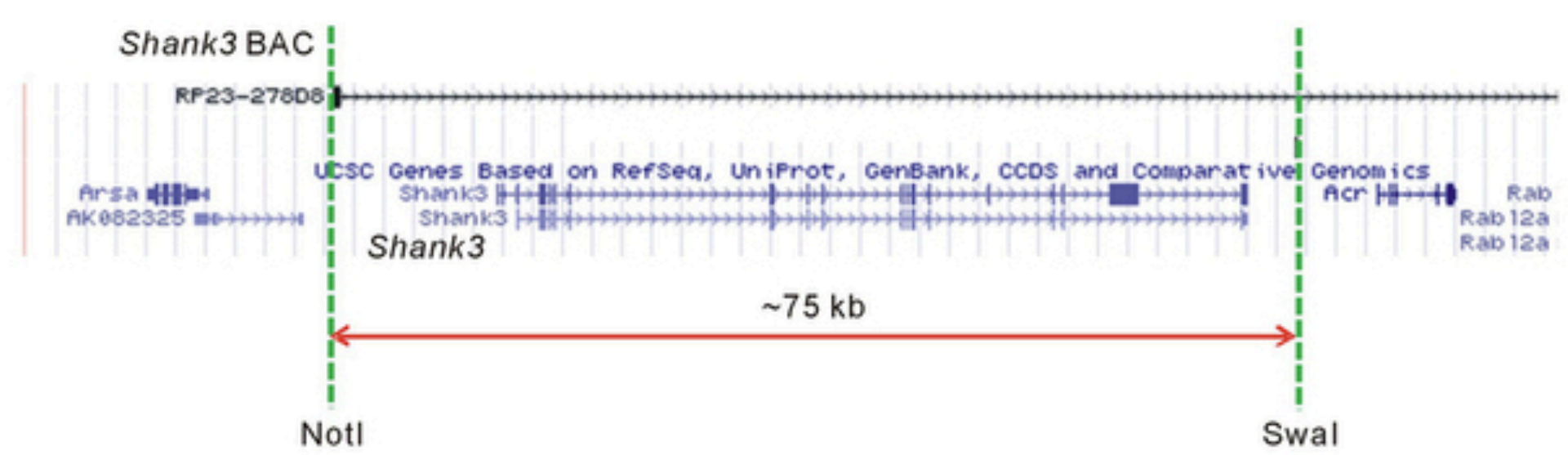
Minimum metadata

life sciences

- Title
- Description
- Creator
- Publication Date
- **Topics**
- **Audience**



Minimum metadata about files?



- Suggestions of other available databases
- Tools to do expression analysis
- Courses on database analysis
- Recent publications involving SHANK3

We would like to

- Define Life Sciences metadata
  - Community based
  - Metadata for EUDAT and all other services in Life Sciences
    - So we can contextualise information between different services

SASI

Scientific Announcement Standards Initiative

Consensus standard for automatic announcement exchange among life science organizations

GOBLET

Global Organisation for Bioinformatics Learning, Education & Training

Bioinformatics Training Network

SoIBio

SGBC

APBioNET

SEB Society for Experimental Biology

biosharing Run By Oxford e-Research Centre

Fundação Calouste Gulbenkian Instituto Gulbenkian de Ciência

CSC

itico

nowgen engagement • education • training

SIB Swiss Institute of Bioinformatics

Radboud University Nijmegen Medical Centre Nijmegen Centre for Molecular Life Sciences

bioinformatics.ca

edinburgh genomics

TGAC The Genome Analysis Centre™

nbic netherlands bioinformatics centre

Bits Bioinformatics ITalian Society

EMBL-EBI

CBAT Computational Genomics Analysis and Training

Australian Bioinformatics Network

iSCB International Society for Computational Biology

CPGR Centre for Proteomic & Genomic Research

ASBCB

SeqAhead

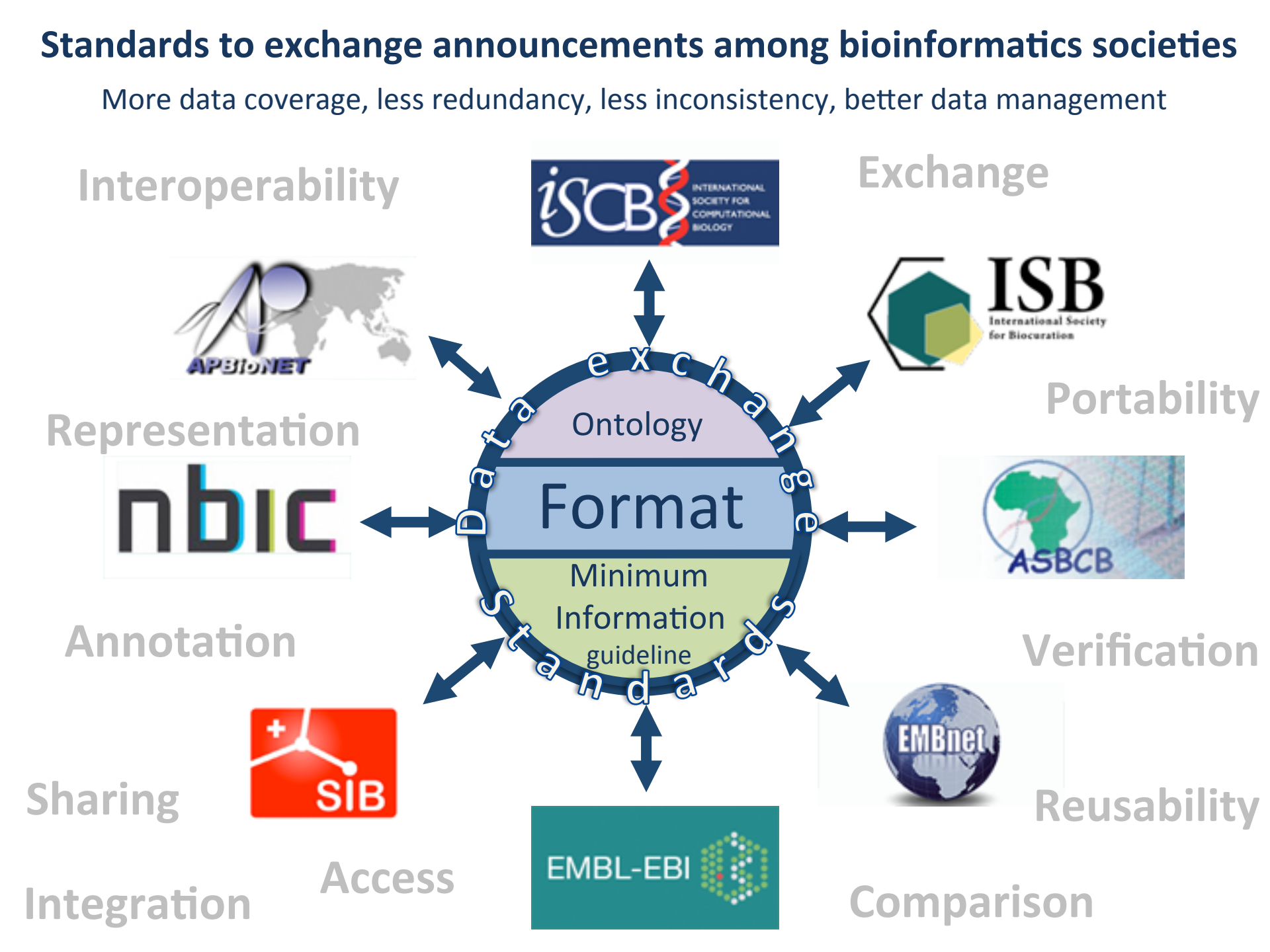
ISB International Society for Biocuration

The Sainsbury Laboratory

EMBnet The Global Bioinformatics Network

Standards to exchange announcements among bioinformatics societies

More data coverage, less redundancy, less inconsistency, better data management



	CATEGORY	TYPE	FIELD	iAnn	EBI	Bioinformatics.c
Minimun	General	text (1)	**Id**	x	i.d. (pri)	x
Minimun	General	text (1)	**Title**	x	Title	x
Recommended	General	text (1)	**Subtitle**	x		
Minimun	General	text (1)	**Description**		Overview	Objectives
Recommended	General	text (1)	**Prerequisites**		Overview/ Prerequisites	
Recommended	General	programme (1)	**Programme**		Programme	
Optional	General	text (1)	**Comments**			x
Minimun	General	fees (n)	**Fees**		Registration Fees	x
Optional	General	discount (1)	**Discount**		Registration	
Recommended	General	text (1)	**Accreditation**			
Optional	General	text (1)	**Status**		Registration	
Recommended	General	text (n)	**Eligibility**		Participation	
Recommended	General	int (1)	**Capacity**		Registration	x
Minimun	People	person (n)	**Contact**		Admin_support	x
Optional	People	person (n)	**Submitter**	x		
Recommended	People	person (n)	**Organizers**		Organisers	
Optional	People	person (n)	**Speakers**		Trainers	
Minimun	Organization	organization (n)	**Host institution**	x	Venue/Overvie	
Optional	Organization	organization (n)	**Sponsor**	x	Sponsors	

You are logged in. Username: **Manny**

Add basic details

Generic

Add more details?

Title	*	Title of the resource
Description	*	
Creator		author
Open Access		ON
Licence		
Keywords		keyword1, keyword2, ...

BBMRI

Study ID	
Study name	
Study Description	
Principal Investigator	
Study design	None selected
Disease	
Categories of data collected	None selected
Planned sampled individuals	
Planned total individuals	
Sex	None selected
Age interval	
Material type	None selected

** indicates required field*

Select files

Start upload

Remove

ics

IN

Biomedical Research

Email1: mc@manuelcorpas.com
Email2: manuelcorpas@tgac.ac.uk
Web: http://goo.gl/ELtV4
Blog: http://manuelcorpas.com
Facebook: http://www.facebook.com/corpasgenome
Twitter: https://twitter.com/manuelcorpas
Linkedin: http://www.linkedin.com/in/manuelcorpas
YouTube: https://www.youtube.com/user/manuelcorpas2
Google+: https://plus.google.com/u/0/101146878928866340609/posts
Google Citations: http://scholar.google.com/citations?user=vBV3CA0AAAAJ

Something additional to minimum metadata

- Standard format
- If we have something that is common for the community, then we can
  - People who search EBI services can see there are things in EUDat that are related
  - EUDat defines metadata, so does figShare, Google, etc. Each organisation defines metadata independently without community agreement