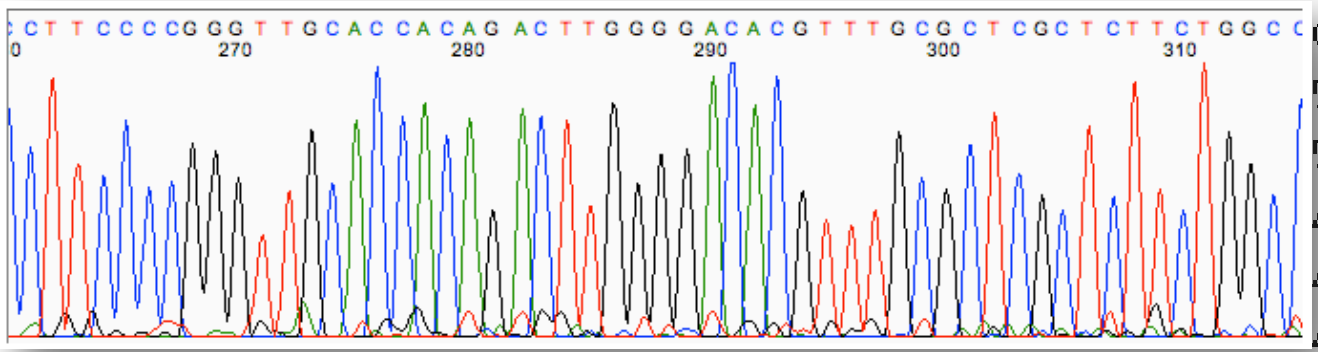
Integrating open articles with open data in the life sciences with Europe PubMed Central

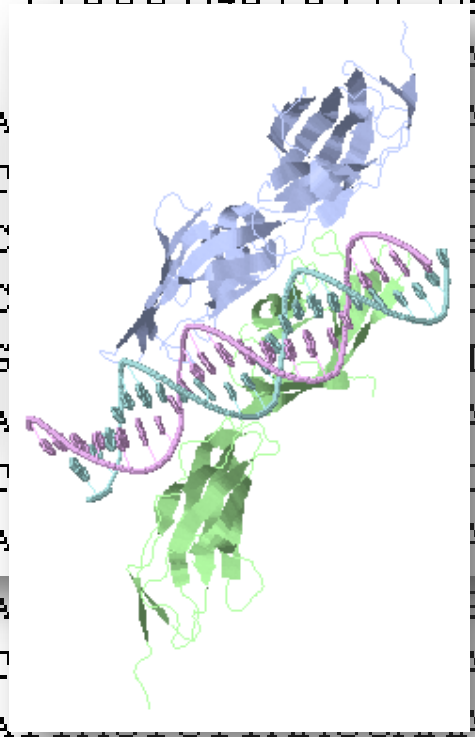
Open Repositories 2014, Helsinki

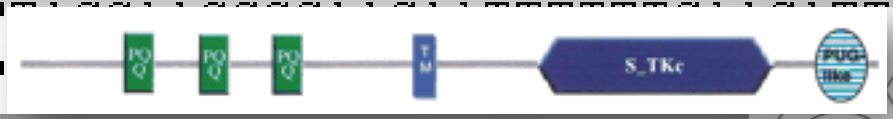
About EMBL-EBI

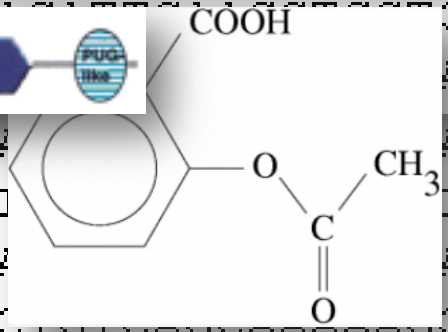
- Part of the European Molecular Biology Laboratory
- International, non-profit research institute
- Europe's hub for biological data services and research

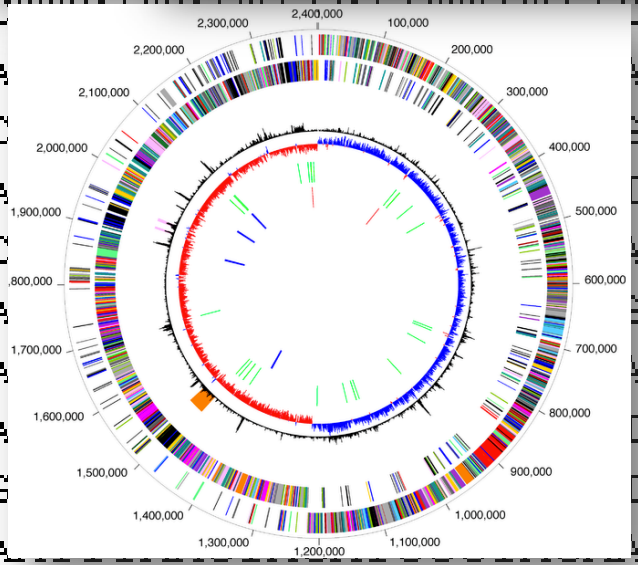
Life Science Data

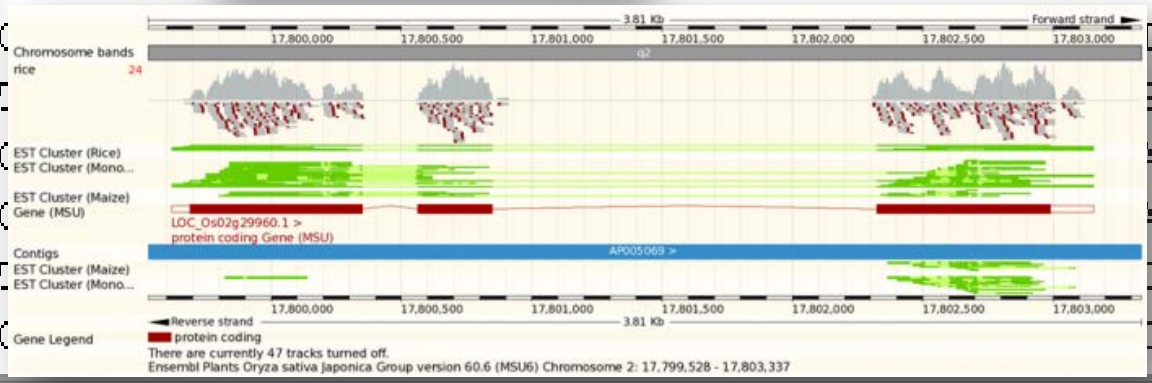












Europe PubMed Central

EMBL-EBI

Journal Data Publishing

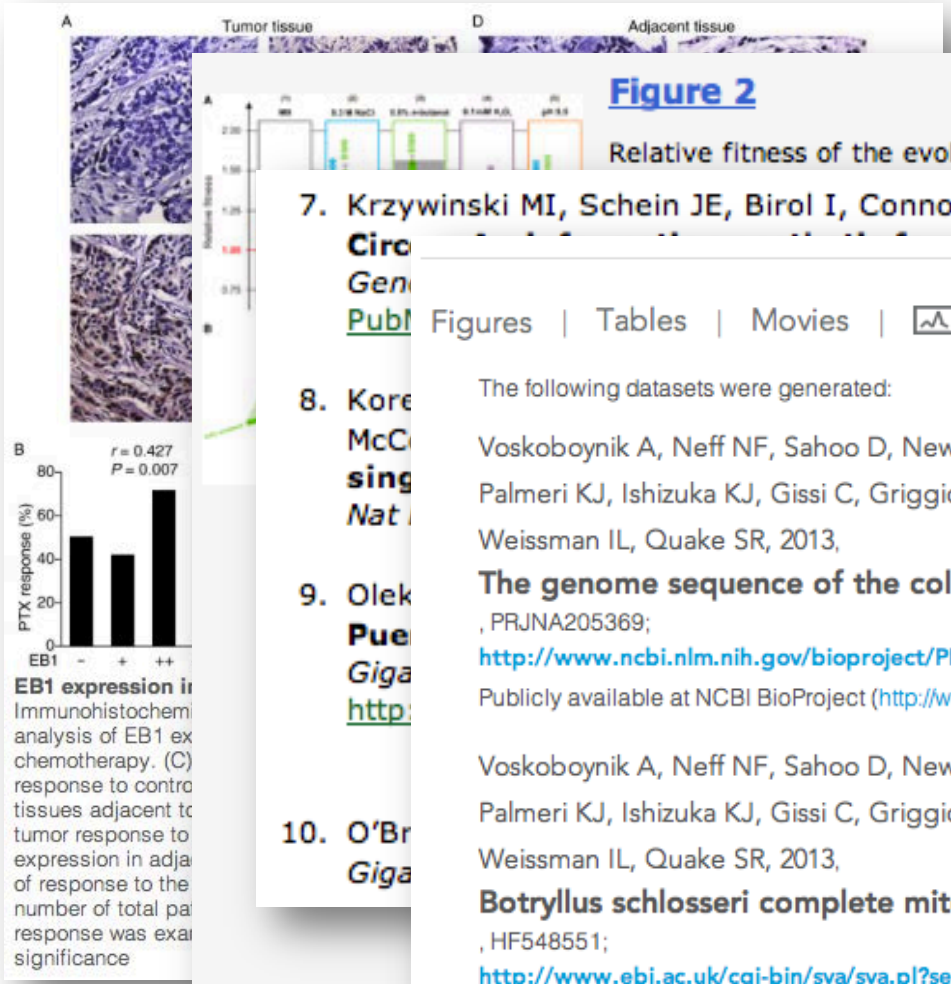


Figure 2

Relative fitness of the evolved populations under all environmental

7. Krzywinski MI, Schein JE, Birol I, Connors J, Gascoyne R, Horsman D, Jones SJ, Marra MA:
Circ
Gen
Publ

8. Kore
McC
sing
Nat

9. Olek
Pue
Giga
http

10. O'Br
Giga

Figures | Tables | Movies | **Major datasets** | Additional files

The following datasets were generated:

Voskoboynik A, Neff NF, Sahoo D, Newman AM, Pushkarev D, Koh W, Passarelli B, Fan CH, Mantalas GL, Palmeri KJ, Ishizuka KJ, Gissi C, Griggio F, Ben-Shlomo R, Corey DM, Penland L, White RAIII, Weissman IL, Quake SR, 2013,

The genome sequence of the colonial chordate, Botryllus schlosseri

, PRJNA205369;

http://www.ncbi.nlm.nih.gov/bioproject/PRJNA205369

Publicly available at NCBI BioProject (http://www.ncbi.nlm.nih.gov/bioproject).

Voskoboynik A, Neff NF, Sahoo D, Newman AM, Pushkarev D, Koh W, Passarelli B, Fan CH, Mantalas GL, Palmeri KJ, Ishizuka KJ, Gissi C, Griggio F, Ben-Shlomo R, Corey DM, Penland L, White RAIII, Weissman IL, Quake SR, 2013,

Botryllus schlosseri complete mitochondrial genome, isolate sc6a-b

, HF548551;

http://www.ebi.ac.uk/cgi-bin/sva/sva.pl?search=Go&query=HF548551

Publicly available at the European Nucleotide Archive (http://www.ebi.ac.uk/ena/).

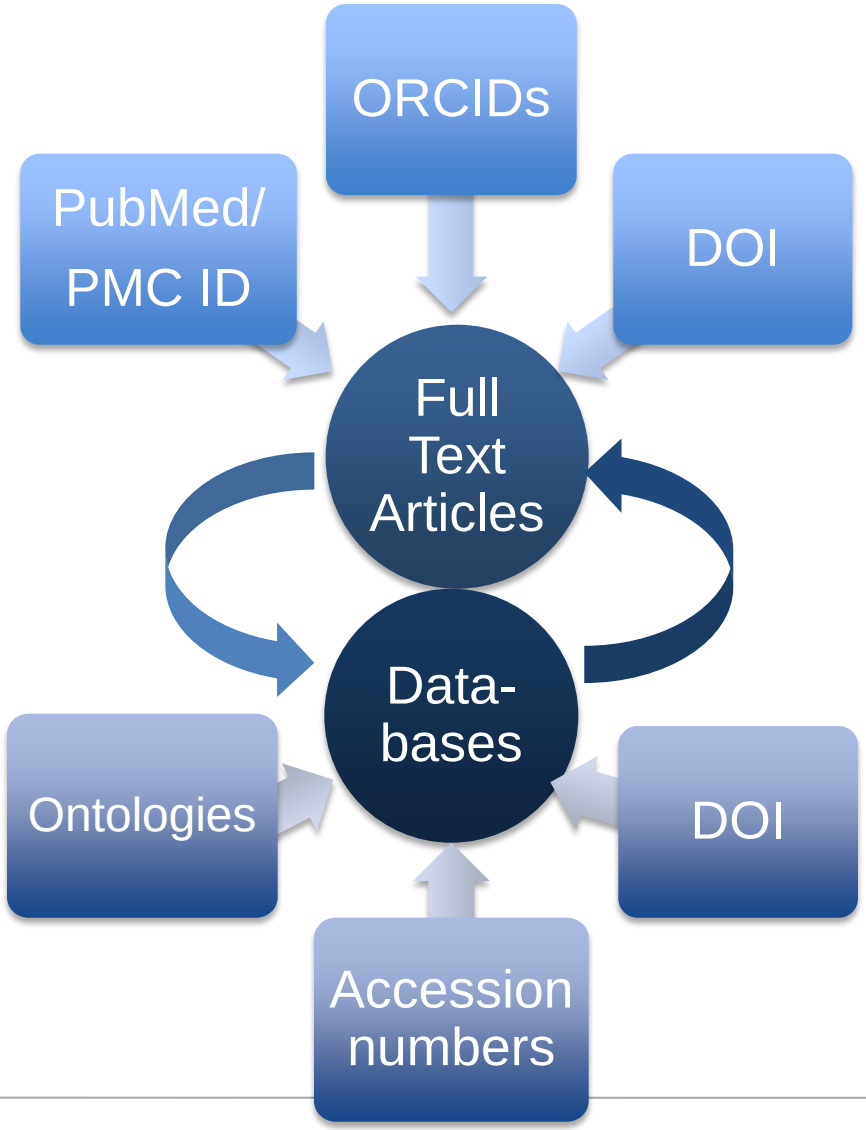
Data Citation Principles

- Data as legitimate, citable products of research
- Attribution and credit
- Cited as evidence for a claim
- Unique identification
- Access
- Persistence
- Specificity and verifiability (provenance)
- Interoperability and flexibility

https://www.force11.org/datacitation

Integrating literature with data

Reason for citation
- Provenance of data
- Data submission statements



Source
- Text mining
- Added by publisher

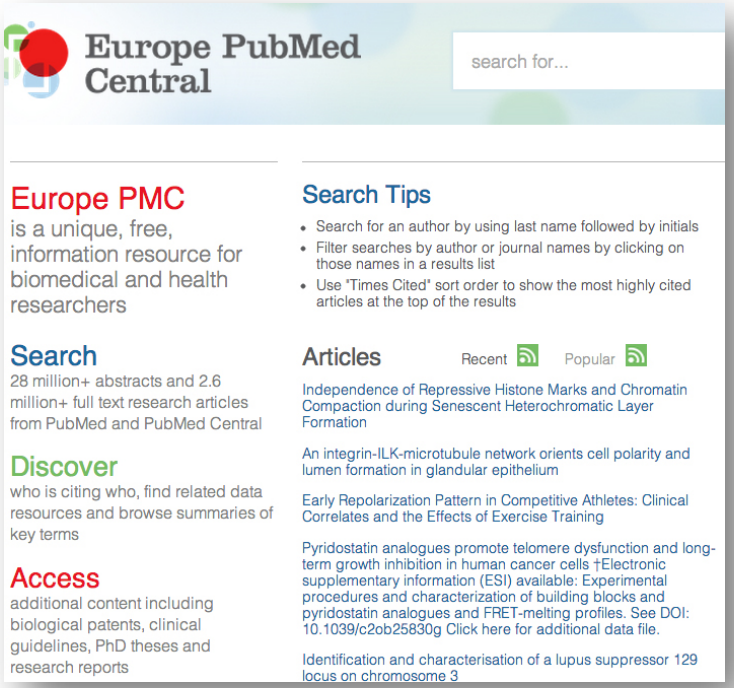
Reason for citation
- Context for data submission
- Curation of data

Source
- Structured meta-data in database record

A stable, freely available, shared repository

Europe PubMed Central:

- Abstracts: 28 million
- Full-text articles: 3 million
- Citation counts, grants, ORCIDs, semantic annotation, data integration
- Grants database
- Labs



Europe PMC is a member of the PMC International Collaboration.

Funded by 26 European funders of life science research

Integrated discovery

Article in Europe PMC

3 Deep links to data

- Sequences
- Structures
- Reactions
- Models
- Prior art

2 Advanced search tools (e.g. citation-count sort)

1 A sustainable, robust repository for
- Individuals
- Funders

4 Article not in Europe PMC

External Links Service

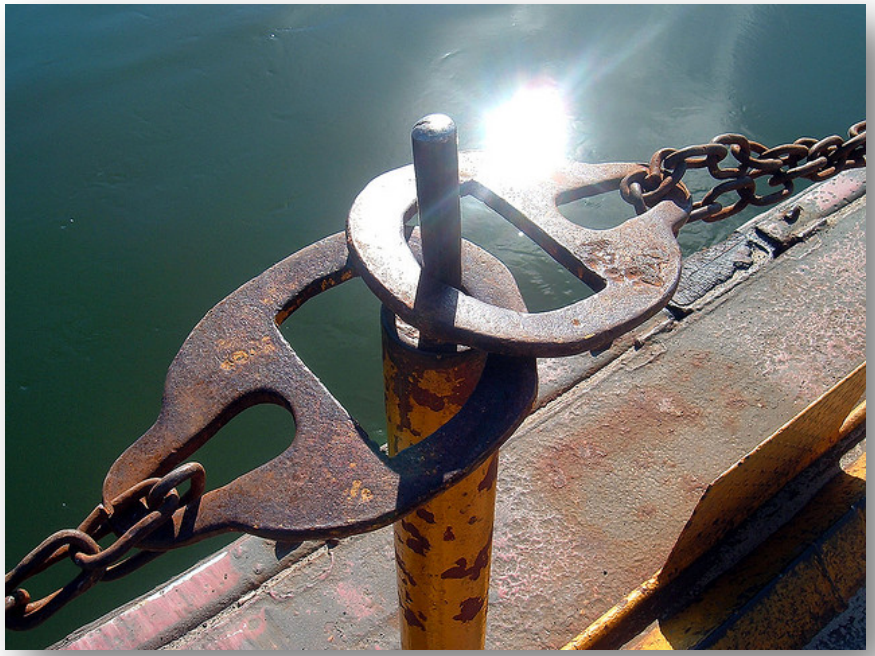


Image courtesy of Diego Torres Silvestre, Flickr

- Links to open, useful content
- Based on identifiers - DOIs
- Increase discoverability
- Links are self administered

- Made available via website and APIs

http://europepmc.org/LabsLink

Hands-On: Interlinking Institutional Repositories and Europe PubMed Central
Najko Jahn, Bielefeld University, Germany, Thursday am, P6A

A network of repositories

Europe PubMed Central

EMBL-EBI

Cross Disciplinary Integration

Prokaryotic community structure and diversity in the sediments of an active submarine mud volcano (Kazan mud volcano, East Mediterranean Sea).
(PMID:20370830)

Abstract | Citations | BioEntities | Related Articles | External Links

Pachiadaki MG, Lykousis V, Stefanou EG, Kormas KA

Environmental Chemical Processes Laboratory, Department of Chemistry Heraklion, Greece.

FEMS Microbiology Ecology [2010, 72(3):429-444]

Type: Journal Article, Research Support, Non-U.S. Gov't
DOI: 10.1111/j.1574-6941.2010.00857.x

Abstract

We investigated 16S rRNA gene diversity at a high sediment depth
cm) in an active site of the Kazan mud volcano, East Mediterranean
374 bacterial clones were analysed, which were attributed to 38 and
respectively (> or = 98% similarity). Most of the archaeal phylotype
-3 members originating from habitats where anaerob
they occurred in sediment layers with no apparent A
Proteobacteria were the most abundant and diverse
dominating in most sediment layers and these were
cycling. The Deltaproteobacteria included several o
of the bacterial phylotypes belonged to 15 known p
representatives from similar habitats. Diversity inde
Archaea and Bacteria, respectively, revealing differ
20 cm below the sea floor, the prokaryotic communi
Archaea and Bacteria. Our study revealed different

Nucleotide Sequences

Found 237 European Nucleotide Archive record(s) citing this article out of which 10 are displayed

- Uncultured archaeon clone KZNMV-10-A9 16S ribosomal RNA gene, partial sequence. (EMBL:FJ712378)
- Uncultured bacterium clone KZNMV-0-B30 16S ribosomal RNA gene, partial sequence. (EMBL:FJ712418)
- Uncultured bacterium clone KZNMV-0-B42 16S ribosomal RNA gene, partial sequence. (EMBL:FJ712426)
- Uncultured bacterium clone KZNMV-0-B56 16S

PANGAEA

Data Publisher for Earth & Environmental Science

- Supplementary material: Pachiadaki, MG et al (2011). Prokaryotic community structure and diversity in the sediment of Kazan mud volcano. PANGAEA
http://dx.doi.org/10.1594/PANGAEA.762956

Europe PubMed Central

EMBL-EBI

Knowledge Integration

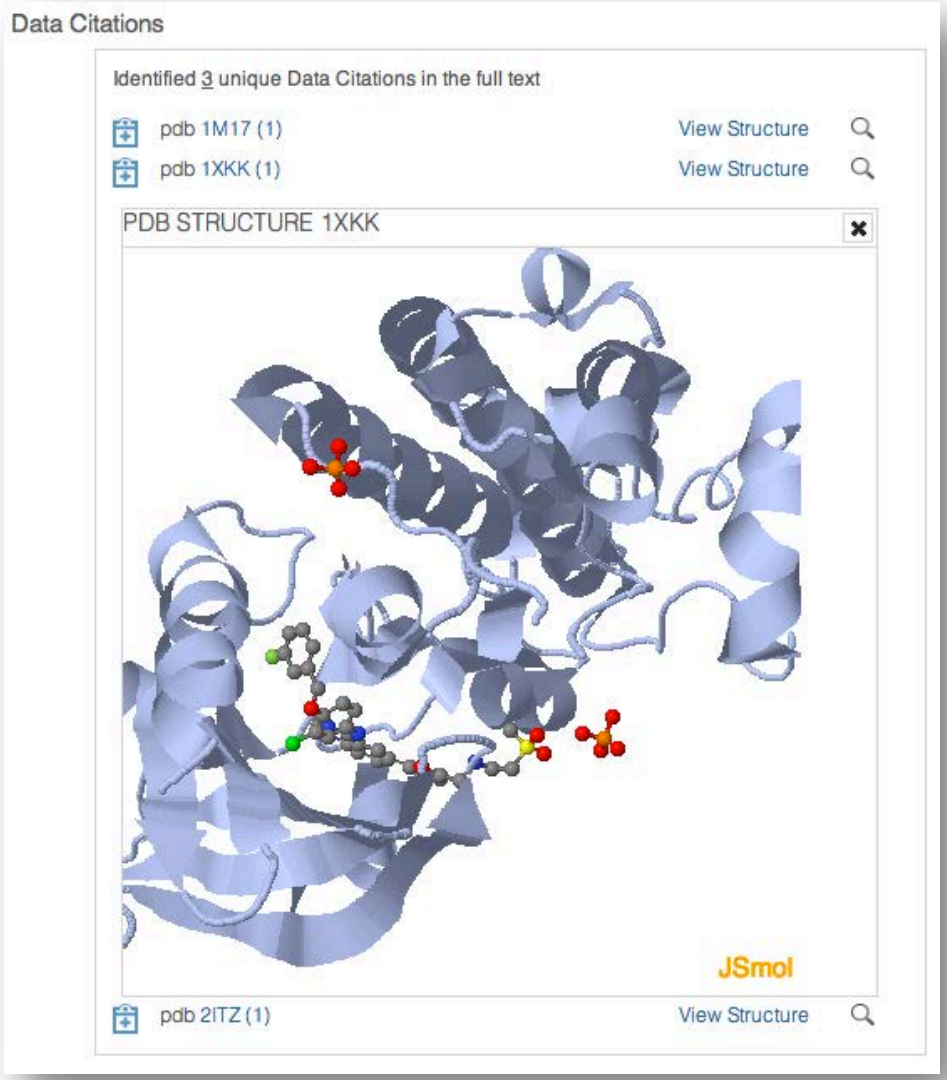
Covalent EGFR inhibitor analysis reveals importance of reversible interactions to potency and mechanisms of drug resistance.

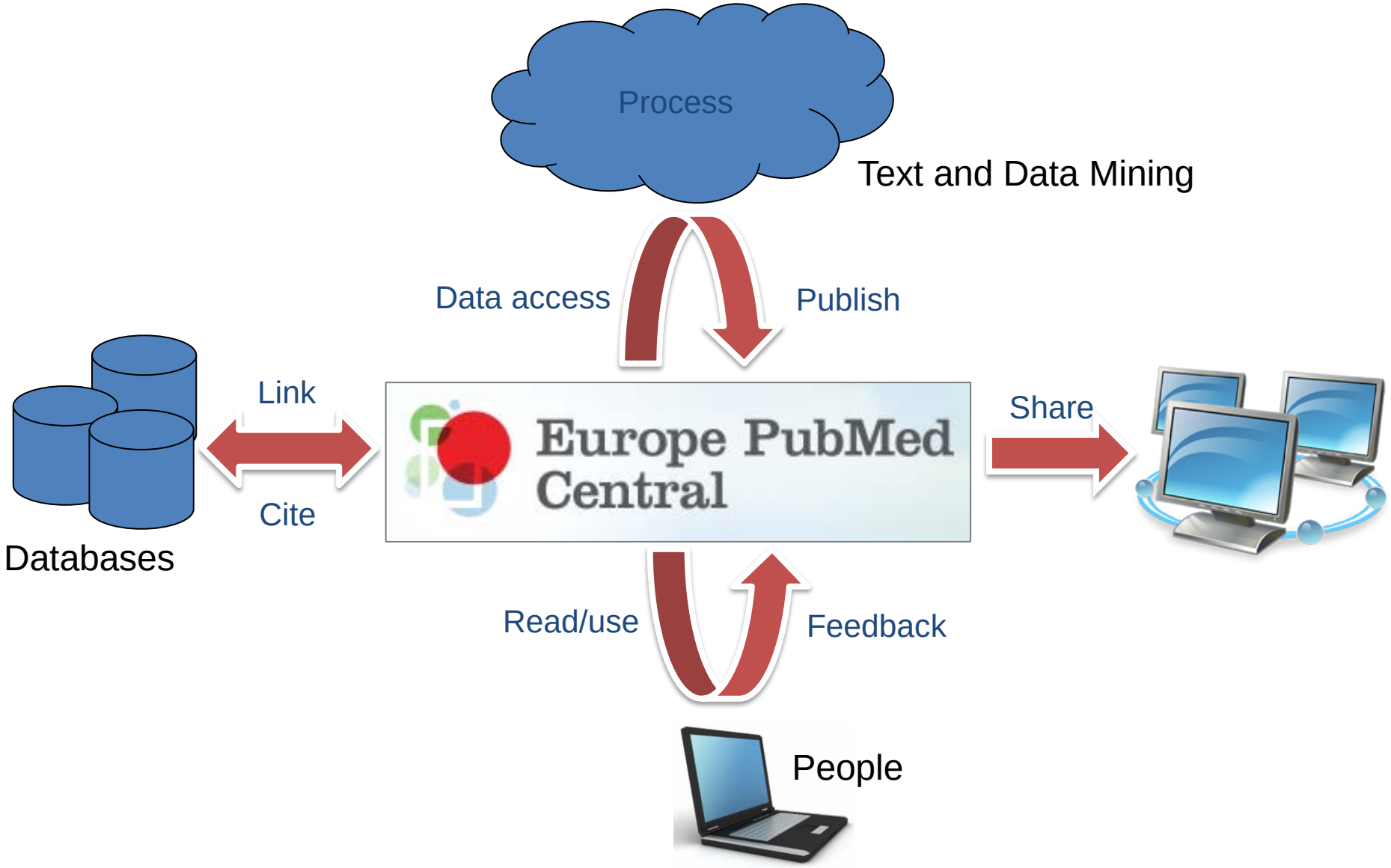
PMC3890870

- Alt Metrics

- + Software =
- Bring data to life within article context

- RESTful APIs




Process
Text and Data Mining
Data access
Publish
Link
Cite
Databases
Europe PubMed Central
Share
Read/use
Feedback
People

What's Missing?

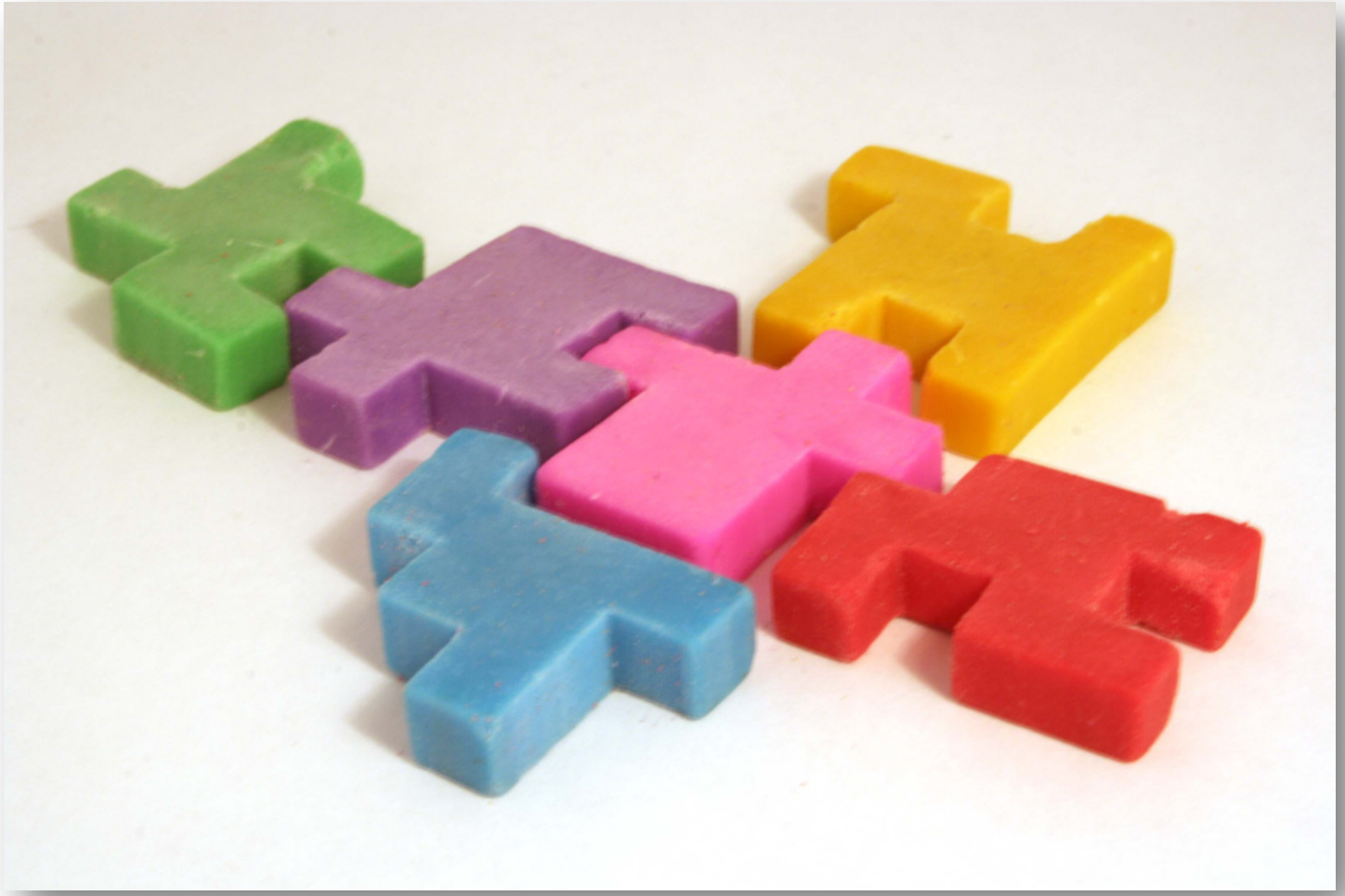
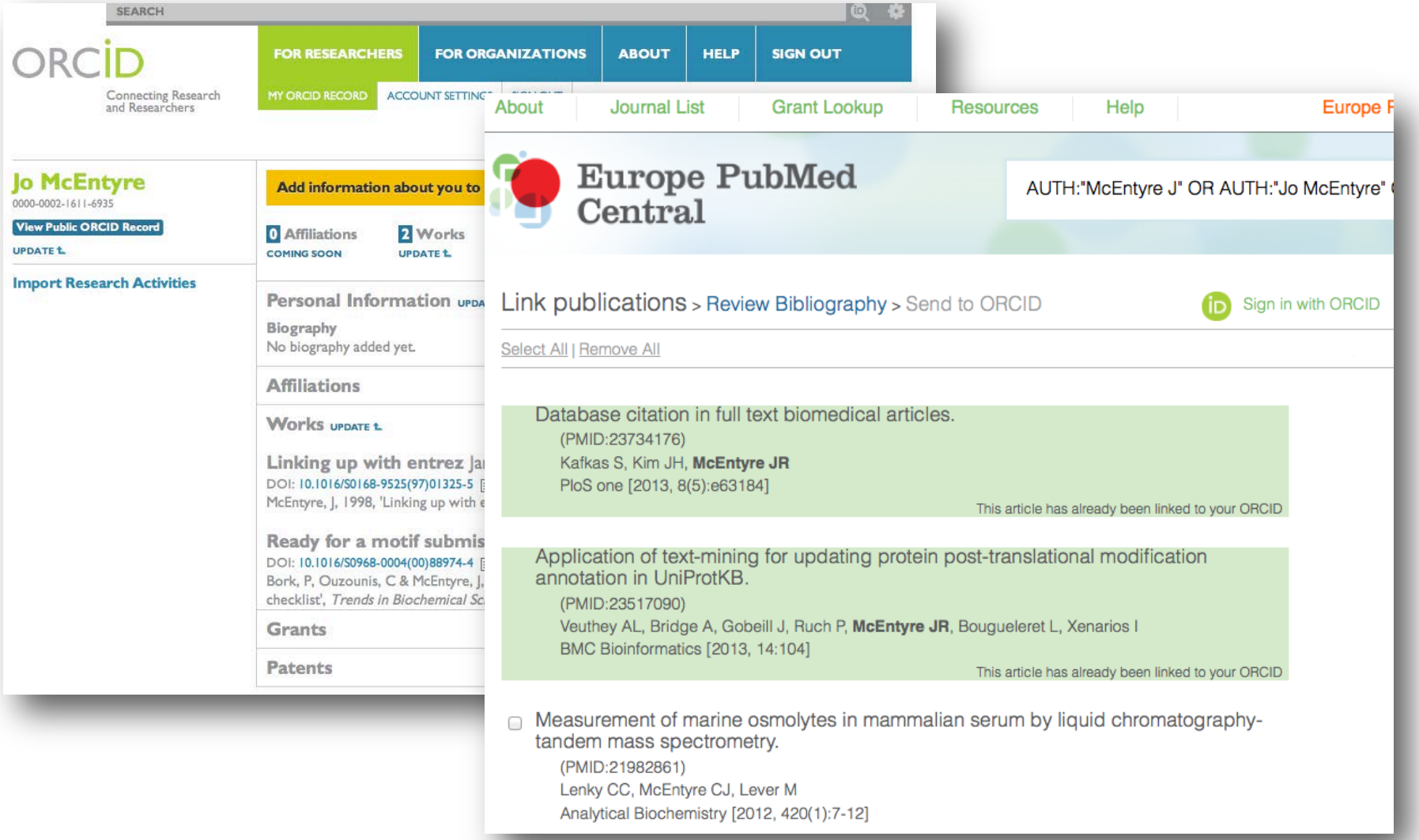


Image by David Singleton, Flickr

Linking works to people



Initial sequencing and analysis of the human genome.

(PMID:11237011)

C, Stange-Thomann N, Stojanovic N, Subramanian A, Wyman D, Rogers J, Sulston J, Ainscough R, Beck S, Bentley D, Burton J, Clee C, Carter N, Coulson A, Deadman R, Deloukas P, Dunham A, Dunham I, Durbin R, French L, Grafham D, Gregory S, Hubbard T, Humphray S, Hunt A, Jones M, Lloyd C, McMurray A, Matthews L,

Abstract

Highlight Terms

☐ Gene Ontology(1) ☐ Species(1)

The human genome holds an extraordinary trove of information about human development, physiology, medicine and evolution. Here we report the results of an international collaboration to produce and make freely available a draft sequence of the human genome. We also present an initial analysis of the data, describing some of the insights that can be gleaned from the sequence.

ORCIDs

0000-0002-3908-1122
0000-0001-9251-070X
0000-0002-1767-9318
0000-0001-7116-6364
0000-0002-6982-4660
0000-0001-8314-8497
0000-0001-5546-9672
0000-0001-8479-0262
0000-0003-1988-5059
0000-0003-2525-5598

Uptake of ORCIDs in the life sciences

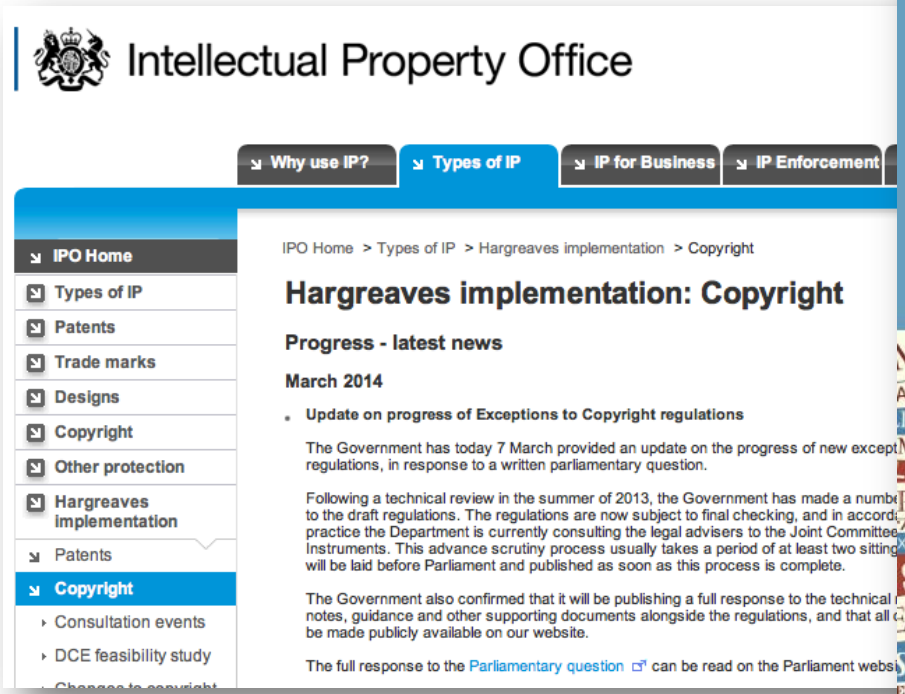
Institutionally, at the EMBL-EBI (June 2nd 2014)

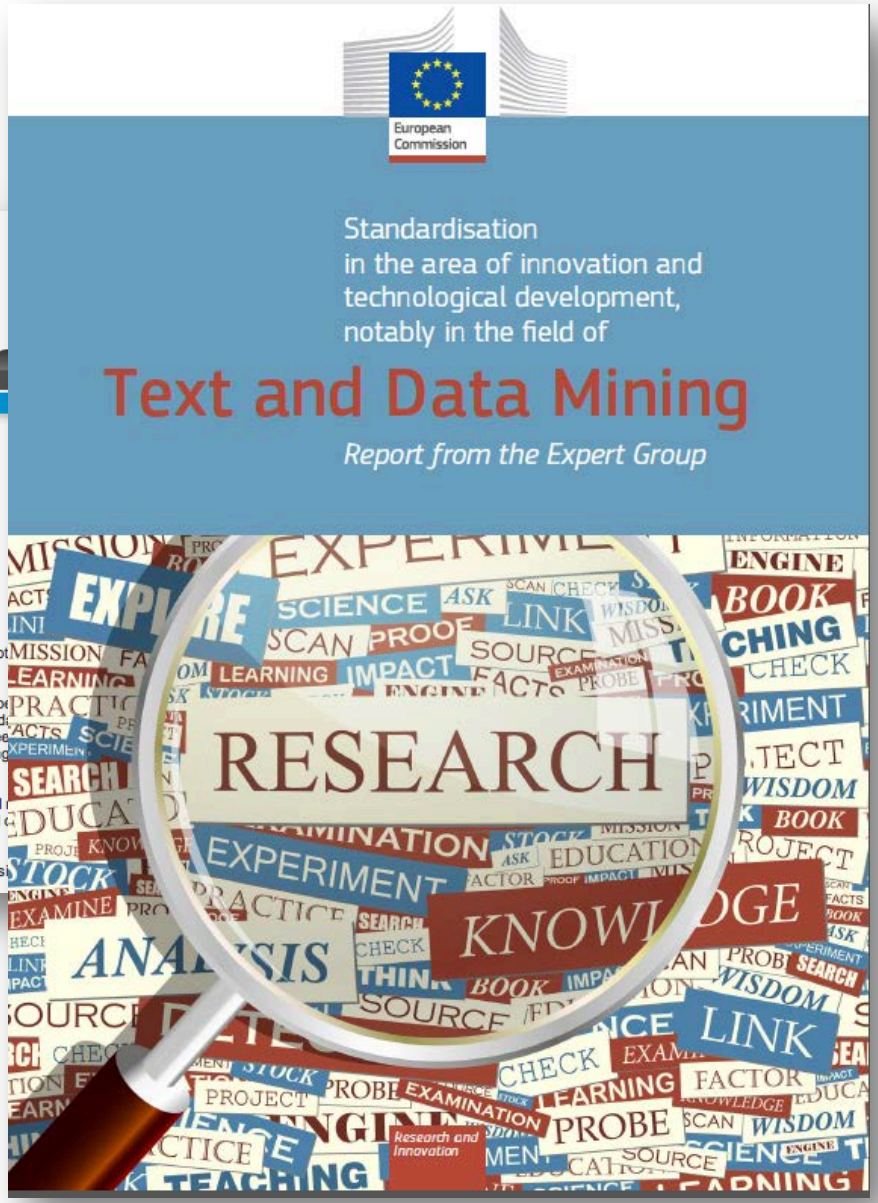
- ~450 ORCIDs have been assigned
- 195 have works attached (~43%)
- 4263 total claims

More widely

- ~ 750,000 articles in Europe PMC have ORCIDs
- This represents ~70K ORCIDs
- ~2/3 of works with DOI (or similar) can be resolved to records in Europe PMC

Copyright Law





Europe PubMed Central

EMBL-EBI

Delivered by: