



**Cloud deployment of FAIR resources:
Success stories from the EOSC-LIFE RI
community**

OmicsDI Cloud

Henning Hermjakob, EMBL-EBI
hhe@ebi.ac.uk



Project description

Develop a cloud-deployed production version of the Omics Discovery Index (OmicsDI), to support efficient integration with cloud-based workflows. In addition, we propose to enable local satellite deployments, to demonstrate OmicsDI use as a community coordination tool for the European and international systems biology community, and to provide API access to OmicsDI impact metrics.

- D 1.1: Production deployment of OmicsDI on EOSC cloud infrastructure.
- D1.2: Open source provision of OmicsDI for local satellite deployment, with optional synchronisation with the OmicsDI public instance.
- D 2.1: Integration of FAIRDOMhub and Virtual Parts Repository into the OmicsDI public cloud deployable instance (D1.1).
- D2.2: BioHackathon, COMBINE.org (or similar event 2021) project for inclusion of additional systems and synthetic biology resources into OmicsDI.
- D 2.3: Infrastructure (API and Web) to enable grant agencies and journals to retrieve the impact metrics of each dataset deposited in OmicsDI. Similar to Altmetrics, this will allow to display a version of the current OmicsDI “rosette” on external websites, or retrieve the data for further external processing.





Organism, repository, gene, tissue, accession

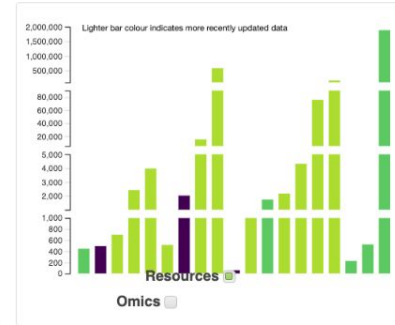
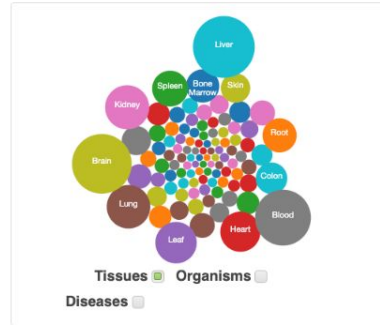
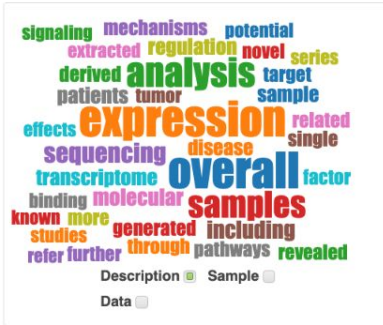
Advanced Search

Examples: Cancer, Homo sapiens, Orbitrap, O9HAU5, Phospho, HeLa, PXD001416

Omics Discovery Index

Consistent dataset discovery across 23 repositories from multiple omics types

D1.1 still in progress: Kubernetes deployment currently in refactoring



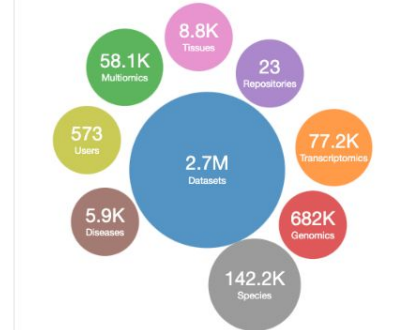
Latest Datasets

- Jun 6 '22 FILIP1 CT pull-down
- May 20 '22 Proteomic Landscape of Pancreatic C...
- May 20 '22 Proteome of CBR1 transgenic SK-OV...
- May 20 '22 Comparative Metabolomic Profiling of...
- May 19 '22 Extensive tumour profiling in primary ...
- May 19 '22 Kim2011_VvuMBEL943_2022
- May 19 '22 Novak2022 - Mitotic kinase oscillation
- May 19 '22 MS analysis of phosphopeptides enri...
- May 19 '22 A non-canonical vitamin K cycle is a ...
- May 19 '22 Human plasma metabolites of SCAP1...
- May 18 '22 Lopez2022 - Genome-scale metaboli...
- May 18 '22 Nikolov2020 - p53-miR34 model

Most Accessed Datasets

- 4112 E-MTAB-2037
- 10737 Kholodenko1999 - EGFR signaling
- 4442 Transcription profiling of Streptococcus go...
- 4441 Identification of proteins interacting with In...
- 4264 WTCCC case-control study for Bipolar Dis...
- 4171 58C
- 4853 A metabolomic study of urinary changes i...
- 3808 Cytokine interactome alterations promote ...
- 3503 Iridovirus and Microsporidian Linked to Ho...
- 3420 M.musculus - Heart, Protein peptide coun...
- 3205 RNA-seq time course analysis of human a...
- 3204 M.musculus - Saliva secreting gland, Prot...

Statistics



Satellite Deployment: ModelXchange



Created satellite instance for systems biology (D1.2).

Included three new repositories (PMR, FAIRDOMHub, Cell Collective) (D2.1)(D2.2).

Supports cross-repository search in fragmented systems biology domain, increasing FAIRness.

The screenshot shows the ModelXchange website interface. At the top, there is a navigation bar with links for 'ModelXchange', 'Browse', 'Submit Data', 'Databases', 'API', and 'Help'. Below the navigation bar is a search bar with the placeholder text 'Organism, repository, gene, tissue, accession' and a search button labeled 'Advanced Search'. The main content area is divided into several sections:

- Description:** A word cloud containing terms like 'cc0', 'please', 'syst', 'models', 'use', 'refer', 'more', 'ii', 'cite', 'biol', 'information', 'biomodels', 'law', 'database', and 'kinetic'.
- Organisms:** A bubble chart showing the distribution of data across various organisms, with 'Homo sapiens' being the largest category. Other categories include 'Mus musculus', 'cellular organisms', 'Bacteria', 'Mammals', 'plant', 'plantae', 'Soil', and 'Saccharomyces cerevisiae'.
- Omics:** A bar chart showing the number of resources for different omics categories. The y-axis ranges from 0 to 2,000,000. The x-axis is labeled 'Resources' and 'Omics'.
- Latest Datasets:** A list of recent datasets with dates and titles, such as 'May 19 '22 Kim2011_VvuMBEL943_2022' and 'May 19 '22 Novak2022 - Mitotic kinase oscillation'.
- Most Accessed Datasets:** A list of frequently accessed datasets, such as '10737 Kholodenko1999 - EGFR signaling' and '2421 Swainston2016 - Reconstruction of human meta...'.
- Repository:** A section titled 'Find your Repository' with a search box and a list of repositories: 'BioModels (2415)', 'PMR (688)', 'FAIRDOMHub (387)', and 'Cell Collective (225)'.



Redundancy detection

Based on shared publication reference, we can identify systems biology models closely related or duplicated across multiple repositories, and reference them.

ModelXchange consortium aims to avoid redundant curation.

ModelXchange Browse Submit Data Databases API Help - omics_type: "RecuratedModel" AND repository: "BioModels" - Advanced

SB Dataset Information

Flis2015 - Plant clock gene circuit (P2011.2.1 PLM_71 ver 2) Ontology highlight

ABSTRACT: cL_m_degr, param m1, modified to ensure light rate > dark rate. Parameter set from PLM_67v2_LDLLLLDs_newFFT_1, with modification to m1 (= old_m1 - m2).

OTHER RELATED OMICS DATASETS IN: [654](#) [655](#) [243](#) [314](#) [380](#) [381](#) [388](#) [BIOMD0000000597](#)

SUBMITTER: [Andrew J Millar](#)

PROVIDER: [BIOMD0000000598](#) | [BioModels](#) | 2015-11-09

REPOSITORIES: BioModels

[ACCESS DATA](#) [Cite](#) [json](#) [Xml](#)

Dataset's files

Filter Source: primary

<input type="checkbox"/>	Name	Type	Action	DRS
<input type="checkbox"/>	BIOMD0000000598?filename=BIOMD0000000598-biopax2.owl	Owl		
<input type="checkbox"/>	BIOMD0000000598?filename=BIOMD0000000598-biopax3.owl	Owl		
<input type="checkbox"/>	BIOMD0000000598?filename=BIOMD0000000598.m	Other		
<input type="checkbox"/>	BIOMD0000000598?filename=BIOMD0000000598.pdf	Pdf		
<input type="checkbox"/>	BIOMD0000000598?filename=BIOMD0000000598.png	Other		

[Upload file\(s\) to Galaxy server](#) Items per page: 5 1 - 5 of 11 [|<](#) [<](#) [>](#) [>|](#)

Publications

Defining the robust behaviour of the plant clock gene circuit with absolute RNA timeseries and open infrastructure.

Flis Anna A Fernández Aurora Piñas AP Zielinski Tomasz T Mengin Virginie V Suljice Ronan R Stratford Kevin K Hume

Similar Datasets

- [Flis2015 - Plant clock gene circuit \(P2011.5.1 PLM_1043 ver 1\)](#) 2015-11-20 | MODEL1510190004 | BioModels
- [Flis2015 - Plant clock gene circuit \(P2011.4.1 PLM_1042 ver 1\)](#) 2015-11-20 | MODEL1510190003 | BioModels
- [Flis2015 - Plant clock gene circuit \(P2011.6.1 PLM_1044 ver 1\)](#) 2015-11-20 | MODEL1510190005 | BioModels
- [Flis2015 - Plant clock gene circuit \(P2011.3.1 PLM_1041 ver 1\)](#) 2015-11-20 | MODEL1510190002 | BioModels
- [Flis2015 - Plant clock gene circuit \(P2011.1.2 PLM_71 ver 1\)](#) 2015-11-09 | BIOMD0000000597 | BioModels
- [Almualem2020 - Virus-macrophage-tumour interactions in oncolytic viral therapies](#) 2021-08-19 | BIOMD000001033 | BioModels
- [Plant, 1981](#) 2013-11-26 | exposure__f6c9c998e7deb7e4a45c8aa769f7f | Physiome Model Repository
- [Karapetyan2016 - Genetic oscillator network - Repressor Titration Circuit \(RT\)](#) 2015-01-26 | BIOMD0000000587 | BioModels
- [Karapetyan2016 - Genetic oscillator network - Activator Titration Circuit \(AT\)](#)



Impact Metrics

OmicDI provides prototype impact metrics for data based on views, downloads, citations, and references among data objects.

Developed API to allow third parties to use these metrics in their sites (D2.3).

Deployed by BioModels and PRIDE resources.

Potential for use by authors, organisations, funders, etc.

The screenshot shows the ModelXchange Repository interface. On the left, there are search filters for Repository, Disease, Publication Date, UNIPROT ID/AC, ENSEMBL ID, and CHEBI ID. The main area displays a list of models. A red box highlights the impact metrics for the model 'Puri2010 - Mathematical model of the Pathogenesis of Alzheimer's Disease'.

Repository	Disease	Publication Date	UNIPROT ID/AC	ENSEMBL ID	CHEBI ID	Impact Metrics
2017-05-18 MODEL1705030000 BioModels						66 Cite
DallePezze2016 - Activation of AMPK and mTOR by amino acids (Model 2)		2017-05-18 MODEL1705030001 BioModels				66 Cite
Kyrtsos2011 - A systems biology model for Alzheimer's disease (Cholesterol in AD)		2016-05-09 MODEL1504240000 BioModels				66 Cite
Puri2010 - Mathematical model of the Pathogenesis of Alzheimer's Disease	Alzheimer's Disease	2015-04-15 MODEL1400020000 BioModels				212 Views 10 Connections 1 Citations 0 Reanalyses 0 Downloads OmicDI score: 67
Priebe1998 - Ventricular fibrillation						63 Cite



Impact Metrics

OmicDI provides prototype impact metrics for data based on views, downloads, citations, and references among data objects.

Developed API to allow third parties to use these metrics in their sites (D2.3).

Deployed by BioModels and PRIDE resources.

Potential for use by authors, organisations, funders, etc.



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824087

The screenshot shows the PRIDE Proteomics IDentifications Database interface. At the top, the PRIDE logo and navigation menu are visible. The main content area displays the project details for PXD000004, including a summary, description, and publication information. The summary section is expanded, showing the title, description, and contact information. The description details the experimental workflow, from sample preparation to data processing. The publication information includes the authors, title, and journal details. The right sidebar contains various metadata sections such as Properties, Similar Studies, and Project Files. The Project Files section includes a table with columns for Name, Type, Size (M), and Download.

PRIDE
Proteomics IDentifications Database

Home Resources Tools Docs License About Contact

Omics score: 1017

3276 Views
3 Connections
1 Citations
1 Reanalysis
190 Downloads

Project PXD000004

Summary [Identification Results](#)

Title
Identification and quantification of human postmortem frontal cortex proteome with a SILAM mouse brain standard

Description
Prefrontal cortex tissue slices from a cognitively and neuropathologically normal human subject were obtained from the University of Pennsylvania (UPenn) Brain Bank. 300-350 mg grey matter was homogenized in 1.5 ml solution A (0.32 M sucrose, 1mM MgCl₂ and 0.1mM CaCl₂) with a Teflon pestle. ~100 % of the homogenate was saved, solubilized with 1% SDS and clarified by centrifugation. To prepare th...

[Read more](#)

Sample Processing Protocol
See details in reference(s) [22942359](#)

Data Processing Protocol
See details in reference(s) [22942359](#)

Contact
[Matthew MacDonald](#), Psychiatry

Submission Date
26/06/2012

Publication Date
10/12/2012

Publication

MacDonald M, Cicimaro E, Prakash A, Banerjee A, Senholzer SH, Blar IA, Hahn CG. Biochemical fractionation and stable isotope dilution liquid chromatography-mass spectrometry for targeted and microdomain-specific protein quantification in human postmortem brain tissue. *Mol Cell Proteomics*. 2012 Dec; 11, 12: 1670-81., PubMed [22942359](#)

Project Files [Project FTP](#)

Name	Type	Size (M)	Download
PRIDE_Exp_Com...	RESULT	81	FTP
PRIDE_Exp_Com...	OTHER	20	FTP
PRIDE_Exp_Com...	PEAK	43	FTP
PRIDE_Exp_Com...	RESULT	106	FTP
PRIDE_Exp_Com...	OTHER	28	FTP
PRIDE_Exp_Com...	PEAK	57	FTP

Properties

Organism
Homo sapiens (human)

Organism part
Brain

Diseases
Unknown

Modification
monohydroxylated residue
SILAC labeled residue
iodoacetamide derivatized residue

Instrument
instrument model
Q Exactive

Software
Proteome Discoverer 1.3.0.339

Experiment Type
Unknown

Quantification
Unknown

Dataset reuses
1

License
EBI terms of use

Similar Studies

[Rat cardiac mitochondrial proteome and phosphoproteome](#)
2013-07-22

[Phosphoproteomic characterization of influenza A virus infected human macrophages](#)
2016-05-25

[Genetics and proteomics of bovine](#)
2009-08-26

[Unique insights in cervicovaginal Lactobacillus flora and L. crispus proteome changes](#)

Impact Metrics

OmicDI provides prototype impact metrics for data based on views, downloads, citations, and references among data objects.

Developed API to allow third parties to use these metrics in their sites (D2.3).

Deployed by BioModels and PRIDE resources.

Potential for use by authors, organisations, funders, etc.



BioModels Individual Models Search...

Home Browse Submit Curation Help About us Contact us Feedback Login Register

Puri2010 - Mathematical Modeling for the Pathogenesis of Alzheimer's Disease

Overview Files History

Model Identifier MODEL1409240001

Short description Puri2010 - Mathematical Modeling for the Pathogenesis of Alzheimer's Disease
Puri2010 - Mathematical Modeling for the Pathogenesis of Alzheimer's Disease

Encoded non-curated model. Issues:

- Confusing replacement of α_{16} when $t = 3$ years
- Confusing 4th rate equation

This model is described in the article:

Mathematical modeling for the pathogenesis of Alzheimer's disease.
Puri JK, Li L.
PLOS ONE 2010; 5(12): e15176

Abstract:
Despite extensive research, the pathogenesis of neurodegenerative Alzheimer's disease (AD) still eludes our comprehension. This is largely due to complex and dynamic cross-talks that occur among multiple cell types throughout the aging process. We present a mathematical model that helps define critical components of AD pathogenesis based on differential rate equations that represent the known cross-talks involving microglia, astroglia, neurons, and amyloid- β (A β). We demonstrate that the inflammatory activation of microglia serves as a key node for progressive neurodegeneration. Our analysis reveals that targeting microglia may hold potential promise in the prevention and treatment of AD.

This model is hosted on [BioModels Database](#) and identified by: MODEL1409240001.

To cite BioModels Database, please use: [BioModels Database: An enhanced, curated and annotated resource for published quantitative kinetic models.](#)

To the extent possible under law, all copyright and related or neighbouring rights to this encoded model have been dedicated to the public domain worldwide. Please refer to [CC0 Public Domain Dedication](#) for more information.

Formal Related SBML (L2V4)

Mathematical modeling for the pathogenesis of Alzheimer's disease.

Metadata information

is	BioModels Database MODEL1409240001
isDescribedBy	PubMed 21179474
hasTaxon	Taxonomy Homo sapiens
isVersionOf	Gene Ontology positive regulation of inflammatory response Gene Ontology positive regulation of microglia differentiation
hasProperty	Human Disease Ontology Alzheimer's disease Mathematical Modelling Ontology Ordinary differential equation model

Curation status Non-curated

Modelling approach(es) ordinary differential equation model

Tags

Connected external resources

OmicDI Impact Metrics

Rosette chart is a feature that shows multiple different normalised metrics of this dataset, including Reanalysis, Citations, Views, Downloads and Connections. Click here to see the Rosette chart of this model via OmicDI service.

Back to top of page



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824087



Impact of the Project on the Research Infrastructure(s)

- Limited technical impact:
 - Implemented “standard” Kubernetes deployment on EMBL-EBI embassy cloud
- Significant community impact:
 - Specialisation of OmicsDI to ModelXchange may have pilot impact for other fields.
 - ModelXchange is expected to have a significant impact on the systems biology community, enabling cross-repository metadata-based discovery.
 - A previous, related effort (ProteomeXchange) contributed to changing data culture in proteomics from “closed” to “open”.
 - Impact metrics for data are a logic follow-up to the FAIR data efforts.
 - Long term indirect impact.





Impact of the Project on EOSC

- Multi-omics, multi-repository metadata discovery and metrics may have pilot function for EOSC.
- Specialisation of multi-omics infrastructure to valuable tool for specific omics field is a “simplified case study” for the EOSC challenge of providing general infrastructures which are useful for specific domains without too much adaptation overhead.





Experience of working in EOSC-LIFE and technical teams

- Less interaction with main project than we could have had.
 - Approached as technical project
 - Regular EOSC-LIFE meetings were “just another meeting”, and so often skipped (by me)
 - To look beyond my own immediate project was challenging as a result.
- Focus on existing EBI infrastructure, rather than “broadening out”
 - Easy, but perhaps not best choice





Future work/Sustainability of the project outcome

- Kubernetes deployment of OmicsDI still ongoing, but will be the long term deployment of choice in a hybrid cloud environment which many organisations use/develop.
- ModelXchange is an informal, but active collaboration of systems biology data resources. The presented work is a prototype that will be sustained and further developed in a collaborative context, and support grant applications.
- Dataset impact metrics are a prototype, supporting efforts to value scientific output as more than just publications. Long term perspectives: Exploratory.





Acknowledgements

- EMBL-EBI:
 - Gaurhari Dass
 - Robert Petryszak
 - Yasset Perez-Riverol
 - Tung Nguyen
 - Helen Parkinson
- Phil Gribbon, ITMP, Germany
- Wolfgang Müller, HITS gGmbH
- Tomas Helikar, University of Nebraska
- David Nickerson, University of Auckland

