



# Research Objects

More than the Sum  
of the Many Parts

Carole Goble

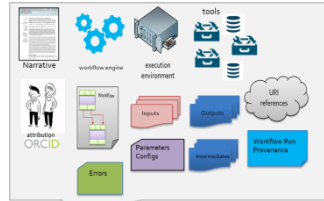
The University of Manchester, UK

EU Infrastructures ELIXIR-UK, FAIRDOM, BioExcel, ISBE ...

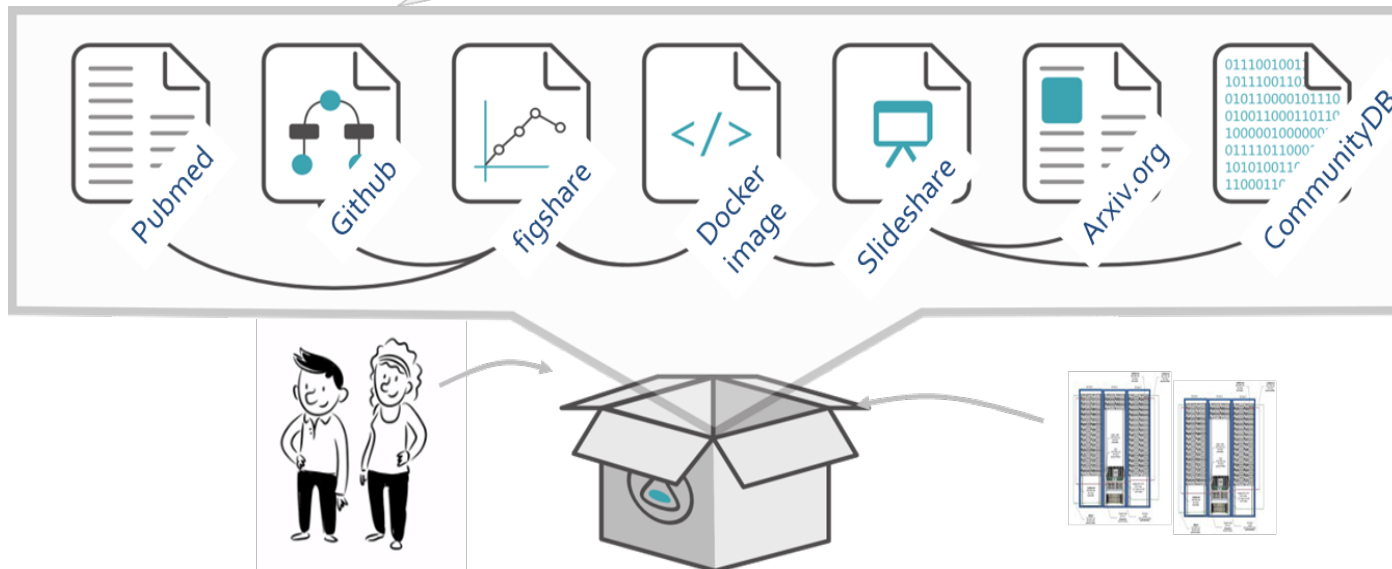
Software Sustainability Institute UK



# Digital Objects Wholes & Parts in an Expanding Ecosystem



**A Digital Package Object Type** composed of many interrelated elements



**A Digital Object** that represents properties in common across all research artefact types, *Common PIDs and Metadata*

**A Digital Package Object** that bundles together and **relates** digital resources of a scientific investigation with context



Nested content  
Heterogeneous elements.  
Distributed and embedded content.  
Externally stewarded content.  
Checklists + Checksums

Citable Reproducible Packaging

# Workflow driven Data Analytics: Research Components are Many and Various

Original Article

## Mapping present and future potential distribution patterns for a meso-grazer guild in the Baltic Sea

Sonja Leidenberger<sup>1,\*</sup>, Renato De Giovanni<sup>2</sup>, Robert Kulawik<sup>3</sup>, Alan R. Williams<sup>4</sup> and Sarah J. Bourlat<sup>5</sup>

Article first published online: 5 SEP 2014  
DOI: 10.1111/jbi.12395

© 2014 The Authors. *Journal of Biogeography*  
Published by John Wiley & Sons Ltd.

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial-NoDerivs License](#), which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

Am score

Additional Information (Show All)

[How to Cite](#) | [Author Information](#) | [Publication History](#) | [Funding Information](#)

[Abstract](#) | [Article](#) | [References](#) | [Supporting Information](#) | [Cited By](#)

[Enhanced Article \(HTML\)](#) | [Get PDF \(1271K\)](#) | [Manchester University](#)

**Keywords:**  
Climate change; Baltic Sea; ecological niche modelling; e-Science; food web; *Fucus radicans*; *Fucus vesiculosus*; *Gasterosteus aculeatus*; *Idotea*; workflows

**Abstract**

**Aim**

The Baltic Sea is one of the world's largest semi-enclosed brackish water bodies characterized by many special features, including endemic species that may be particularly threatened by climate change. We mapped potential distribution patterns under present and future conditions for a community with three trophic levels. We analysed climate-induced changes in the species' distribution patterns and examined possible consequences for the chosen food web.

northern Europe.

SEARCH

In this issue

Advanced > Saved Searches >

ARTICLE TOOLS

- Get PDF (1271K)
- Save to My Profile
- E-mail Link to this Article
- Export Citation for this Article
- Get Citation Alerts
- Request Permissions

Share | [Facebook](#) | [Twitter](#) | [LinkedIn](#) | [Google+](#)

BioVeL | Workflows | Files | Packs

Back to [Experiment](#) | [About BioVeL](#) | [BioVeL Portal](#) | [Biodiversity Catalogue](#)

Log in | Register

Home | Workflows | Bioclim workflow

**Bioclim workflow**

Created: 2013-09-02 13:44:59 | Last updated: 2014-12-04 15:58:08

[Download Workflow](#) | [Open in OnlineHPC](#)

This workflow takes as input a file containing species occurrence points to create a model with the openModeller Web Service using the Bioclim algorithm. Environmental layers and mask are selected during the workflow. Points are filtered so that only environmentally unique points are used to create the model.

You can use this workflow to either filter a set of species occurrence points excluding redundant points with the same environmental conditions (you can get the result in the output "selected\_csv\_points") or to know the environmental ranges for each variable considering all input points. In the latter case, you can inspect the "model\_xml" to find this information in the Maximum and Minimum attributes of the Bioclim element. Values are separated by a space in the same order of the layers. Please note that there are no model tests or projections in this workflow.

For more information about the input file format, please look at the documentation for the corresponding parameter. If you use the default occurrence points you should know that Gammarus tigrinus is an aquatic species, so you need to choose marine environmental layers during the modelling procedure.

Workflow requirements: When running on Taverna workbench, this workflow requires Internet connection and the Taverna interaction plugin installed.

**Preview**

Workflow Inputs: input\_points, use\_mask, select\_csv\_points

Processing Steps: capture, exclude\_masks, get\_environmental\_layers, bioclim\_selected\_points, select\_layers, capture, exclude\_layers, algorithm\_xml, select\_or\_exclude\_mask, no, empty\_string, yes, testing\_script, create\_model, convert\_points\_xml\_to\_csv

Workflow Outputs: layers, model\_xml, model\_img, num\_points, selected\_csv\_points

[Download as scalable diagram \(SVG\)](#)

**Version 4 (latest) (of 4)**

View version: 4 (latest)

**Credits (2)**

(People/Groups)

Renato De Giovanni

BioVeL

**Attributions (0)**

(Workflows/Files)

None

**Tags (13)**

Uploader tags

bio-oracle | bioclim | bioclim | eurostat | filter | harmonized world soil database hwsd | hcat4 | hcat6 | Incofish | openmodeller | points | range | worldcism

[Log in to add Tags](#)

**Shared with Groups (2)**

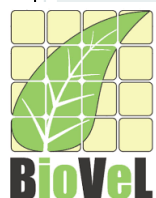
BioVeL

BioVeL Internal

**Featured In Packs (2)**

Mapping present and future predicted distribution patterns for a meso-grazer guild in the Baltic Sea

Application of Ecological Niche Modelling and Earth Observation for the



Added by Biodiversity eLaboratory ... about one hour ago (04/08/14 @ 11:36:52) [more](#)

**File: Fucus vesiculosus current distribution (Biodiversity eLaboratory)**

A low-resolution PNG image of the predicted current niche of Fucus vesiculosus

Added by Biodiversity eLaboratory ... about one hour ago (04/08/14 @ 11:36:28) [more](#)

**File: Fucus vesiculosus current and 2050 difference (Biodiversity eLaboratory)**

A low resolution PNG image for the difference between the current and predicted 2050 niches for Fucus vesiculosus

Added by Biodiversity eLaboratory ... about one hour ago (04/08/14 @ 11:35:41) [more](#)

**File: Fucus vesiculosus occurrences from GBIF (Biodiversity eLaboratory)**

A spreadsheet indicating the source of occurrences of Fucus vesiculosus retrieved from GBIF.

Added by Biodiversity eLaboratory ... about one hour ago (04/08/14 @ 11:25:26) [more](#)

**Workflow: Bioclim workflow with interaction [2] (Renato De Giovanni)**

Taverna 2 BioClim workflow developed as part of the BioVeL project. The workflow filters for environmentally unique points and generates a BioClim model to calculate the environmental range for each variable. The workflow uses the OpenModeller framework

Added by Biodiversity eLaboratory ... 2 hours ago (04/08/14 @ 10:05:40) [more](#)

**Workflow: BioVeL ESW DIFF Basic [2] (Robert Kulawik)**

Taverna 2 ENM Statistical Workflow developed as part of the BioVeL project. The workflow allows the spatial computation of changes in potential distribution maps by comparing the differences between two raster layers using the R statistical environment

Added by Biodiversity eLaboratory ... 2 hours ago (04/08/14 @ 09:52:35) [more](#)

**Workflow: Ecological niche modelling workflow [20] (Renato De Giovanni)**

Taverna 2 Ecological Niche Modelling Workflow developed as part of the BioVeL project. The

Local files

Workflows

## Workflow Commons

Added by Biodiversity eLaboratory ... 4/08/14 @ 10:23:33)

**External: Idotea chelipes - (<http://www.marinespecies.org/aphia.php?p=taxdetails&id=119042>)**

A species of isopod that is prey for Gasterosteus aculeatus and which feeds on the two species of Fucus algae.

Added by Biodiversity eLaboratory ... 2 hours ago (04/08/14 @ 10:22:40)

**External: Idotea balthica - (<http://www.marinespecies.org/aphia.php?p=taxdetails&id=119042>)**

A species of isopod that is prey for Gasterosteus aculeatus and which feeds on the two species of Fucus algae.

Added by Biodiversity eLaboratory ... 2 hours ago (04/08/14 @ 10:21:43)

**External: Fucus radicans - (<http://www.marinespecies.org/aphia.php?p=taxdetails&id=119042>)**

A species of brown algae that is eaten by Idotea

Added by Biodiversity eLaboratory ... 2 hours ago (04/08/14 @ 10:19:43)

**External: Fucus vesiculosus - (<http://www.marinespecies.org/aphia.php?p=taxdetails&id=119042>)**

A seaweed/algae that is eaten by Idotea

Added by Biodiversity eLaboratory ... 2 hours ago (04/08/14 @ 10:17:53)

**External: OpenModeller ecological niche modeling experimentation framework - (<http://openmodeller.sourceforge.net/>)**

Added by Biodiversity eLaboratory ... 2 hours ago (04/08/14 @ 10:00:38)

**Relationships (5)**

1. Fucus vesiculosus 2050 distribution is an output from Ecological niche modelling workflow.
2. Fucus vesiculosus 2050 distribution is an input to BioVeL ESW DIFF Basic.
3. Fucus vesiculosus current distribution is an input to BioVeL ESW DIFF Basic.
4. Fucus vesiculosus current distribution is an output from Ecological niche modelling workflow.
5. Fucus vesiculosus current and 2050 difference is an output from BioVeL ESW DIFF Basic.

**Download**

[Download Pack Items \(ZIP archive\)](#)

10th Anniversary!

Pointer to 3rd Party Data Collection

Pointer to 3rd Party Code

Track?



# Systems and Synthetic Biology: Research Components are Many and Various



## Construction and validation of a detailed kinetic model of glycolysis in *Plasmodium falciparum*

Gerald Penkler<sup>1,2</sup>, Francois du Toit<sup>1</sup>, Waldo Adams<sup>1</sup>, Marina Rautenbach<sup>1</sup>, Daniel C. Palm<sup>1</sup>, David D. van Niekerk<sup>1</sup> and Jacky L. Snoep<sup>1,2,3</sup>

<sup>1</sup> Department of Biochemistry, Stellenbosch University, Matieland, South Africa

<sup>2</sup> Molecular Cell Physiology, Vrije Universiteit Amsterdam, The Netherlands

<sup>3</sup> MB, University of Manchester, UK

### Keywords

enzyme kinetics; glucose metabolism; model workflow; mathematical model; systems biology

### Correspondence

J. L. Snoep, Department of Biochemistry, Stellenbosch University, Private Bag X1, Matieland 7602, South Africa  
Fax: +27 (0)2180 85865  
Tel: +27 (0)2180 85844  
E-mail: jls@sun.ac.za

(Received 19 August 2014; revised 7 February 2015; accepted 13 February 2015)

doi:10.1111/febs.13237

The enzymes in the Embden-Meyerhof-Parnas pathway of *Plasmodium falciparum* trophozoites were kinetically characterized and their integrated activities analyzed in a mathematical model. For validation of the model, we compared model predictions for steady-state fluxes and metabolic concentrations of the hexose phosphates with experimental values for intact parasites. The model, which is completely based on kinetic parameters that were measured for the individual enzymes, gives an accurate prediction of the steady-state fluxes and intermediate concentrations. This is the first detailed kinetic model for glucose metabolism in *P. falciparum*, one of the most prolific malaria-causing protozoa, and the high predictive power of the model makes it a strong tool for future drug target identification studies. The modelling workflow is transparent and reproducible, and completely documented in the SEEK platform, where all experimental data and model files are available for download.

### Database

The mathematical models described in the present study have been submitted to the JWS Online Cellular Systems Modelling Database (<http://jhs.vu.nl/databases/penkler>). The investigation and complete experimental data set is available on SEEK (10.15406/seek.1.investigation.56).

### Introduction

Despite several attempts at a complete eradication of the disease, malaria is still killing more than half a million people per year, mostly small children in sub-Saharan Africa (World Health Organisation Malaria report 2013, [http://www.who.int/malaria/publications/world\\_malaria\\_report\\_2013/en/](http://www.who.int/malaria/publications/world_malaria_report_2013/en/)). The disease is caused by parasitic protozoa of the *Plasmodium* genus, which

have a complicated life cycle consisting of an insect vector and vertebrate host [1]. In the human host, parasite sporozoites first invade liver cells, but the malaria disease symptoms manifest only at a later stage during multiplication of the asexual stages of the parasite in red blood cells (RBCs). The blood life cycle consists of ring, trophozoite and schizont stages, and subsequent

### Abbreviations

3PG, 3-phosphoglycerate; 3PG, 3-phosphoglycerate; ALD, fructose-bisphosphate aldolase; B13PG, 1,3-bisphosphoglycerate; DHAP, glyceraldehyde 3-phosphate; ENO, phosphoenolpyruvate hydratase; F16BP, fructose 1,6-bisphosphate; F6P, fructose 6-phosphate; G3P, glycerol 3-phosphate; G3PDH, glycerol 3-phosphate dehydrogenase; G6P, glucose 6-phosphate; GAP, D-glyceraldehyde 3-phosphate; GAPDH, glyceraldehyde 3-phosphate dehydrogenase; GLC, glucose; GLY, glycerol; HK, hexokinase; LAC, lactate; LDH, lactate dehydrogenase; MCT, monocarboxylate transporter; ODE, ordinary differential equation; PEP, phosphoenolpyruvate; PFK, 6-phosphofructokinase; PGI, glucose 6-phosphate isomerase; PGI, phosphoglycerate kinase; PGM, phosphoglycerate mutase; PK, pyruvate kinase; PYR, pyruvate; RBC, red blood cell; TCA, tricarboxylic acid; TPI, triose-phosphate isomerase.



16 datafiles (kinetic, flux inhibition, runout)



19 models (kinetics, validation)



13 Standard Operating Procedures



3 studies (model analysis, construction, validation)



24 assays/analyses (simulations, model characterisations)

Penkler, G., du Toit, F., Adams, W., Rautenbach, M., Palm, D. C., van Niekerk, D. D. and Snoep, J. L. (2015), Construction and validation of a detailed kinetic model of glycolysis in *Plasmodium falciparum*. FEBS J, 282: 1481–1511. doi:10.1111/febs.13237

Yellow pages

Programmes

Projects

Institutions

People

Experiments

Investigations

Studies

Assays

Assets

Data files

Models

SOPs

Publications

Activities

Presentations

Events

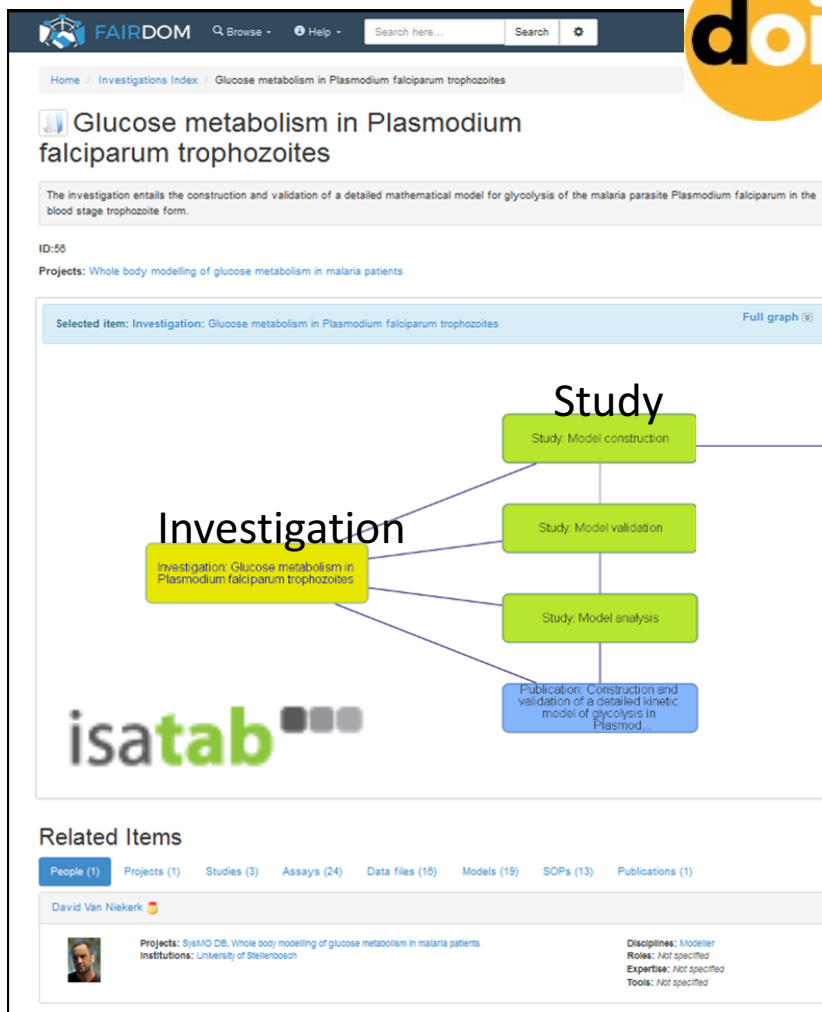
Samples

Sample types

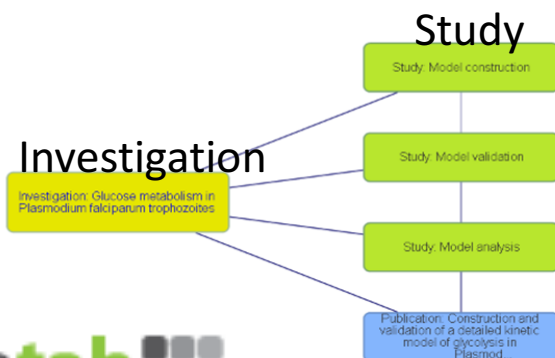
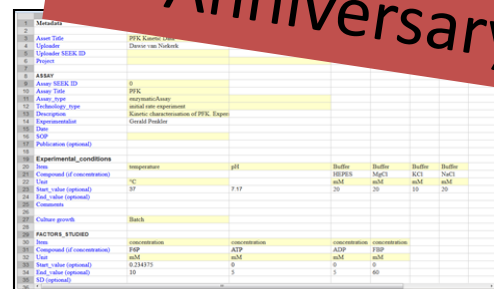
Organisms



9th Anniversary!



The screenshot shows the FAIRDOM interface for the investigation 'Glucose metabolism in Plasmodium falciparum trophozoites'. It includes a search bar, navigation links, and a detailed view of the investigation. The investigation description states: 'The investigation entails the construction and validation of a detailed mathematical model for glycolysis of the malaria parasite Plasmodium falciparum in the blood stage trophozoite form.' Below this, it lists related items: People (1), Projects (1), Studies (3), Assays (24), Data files (16), Models (19), SOPs (13), and Publications (1). The 'Related Items' section shows 'David Van Niekerk' as the project leader, with the project title 'SysMO DB, Whole body modelling of glucose metabolism in malaria patients' and the institution 'University of Stellenbosch'.

Assay	Assay Type	Assay Title	Assay Description	Assay Status	Assay Date	Assay Location	Assay Contact
PFK	Enzyme Assay	PFK Kinetic Assay	Enzyme Assay for PFK Kinetic Characterisation	Completed	2014-05-10	Stellenbosch	David Van Niekerk

Analysis  
Modelling Analysis PFK  
(Assay)

Data

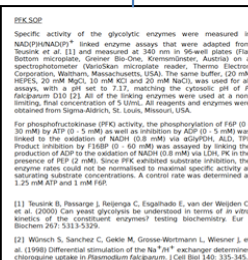
Data file: PFK Kinetic data

Model

Model: PFK Kinetic model

SOP

SOP: PFK Kinetic characterisation



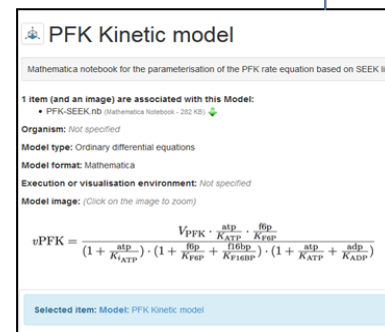
**PFK SOP**

Specific activity of the glycolytic enzymes were measured in  $\mu\text{mol/min/mg protein}$ . Initial enzyme assays that were adopted from Sunkin et al. [1] and measured at 340 nm in 96-well plates (flat bottom microtiter, Greiner Bio-One, Frickenhausen, Austria) on a spectrophotometer (Variankan microplate reader, Thermo Electron Corporation, Waltham, Massachusetts, USA). The same buffer (20 mM HEPES, 20 mM MgCl<sub>2</sub>, 10 mM KCl and 20 mM NaCl) was used for all assays, with a pH set to 7.12, matching the cytosolic pH of *P. falciparum* D10 [2]. All of the linking enzymes were used at a non-limiting, final concentration of 5 U/ml. All reagents and enzymes were obtained from Sigma-Aldrich, St. Louis, Missouri, USA.

For phosphofructokinase (PFK) activity, the phosphorylation of FBP (0 - 10 mM) by ATP (0 - 5 mM) as well as inhibition by ADP (0 - 5 mM) was tested by the addition of NADH (0.5 mM) via ADP:NAD. 5% Product inhibition by FBP (0 - 60 mM) was assessed by titrating the production of ADP to the addition of NADH (0.5 mM) via LNA. PK in the presence of PFK (2 mM). Since PFK exhibited substrate inhibition, the enzyme rates could not be normalised to maximal specific activity at saturating substrate concentrations. A control rate was determined at 1.25 mM ADP and 1 mM FBP.

[1] Sunkin B, Revorge J, Rodriguez C, Engelhardt E, van der Weijden C, et al. (2000) Can yeast glycolysis be understood in terms of *in vitro* kinetics of the constituent enzymatic testing biochemistry. Eur J Biochem 267: 5313-5320.

[2] Wörlich S, Sanchez C, Gekle M, Grosse-Wilde L, Wiesner J, et al. (1998) Differential stimulation of the Na<sup>+</sup>/H<sup>+</sup> exchanger determines intracellular uptake in Plasmodium falciparum. J Cell Biol 140: 335-345.



**PFK Kinetic model**

Mathematica notebook for the parameterisation of the PFK rate equation based on SEEK data

1 item (and an image) are associated with this Model:

- PFK-SEEK.nb (Mathematica Notebook - 202 KB)

Organism: Not specified

Model type: Ordinary differential equations

Model format: Mathematica

Execution or visualisation environment: Not specified

Model image: (Click on the image to zoom)

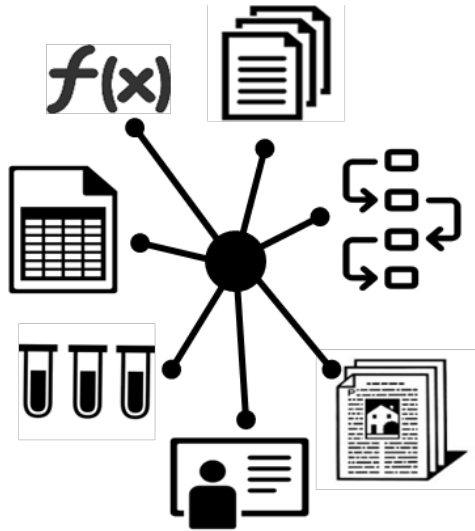
$$v_{\text{PFK}} = \frac{V_{\text{PFK}} \cdot \frac{\text{atp}}{K_{\text{ATP}}} \cdot \frac{\text{fbp}}{K_{\text{FBP}}}}{(1 + \frac{\text{atp}}{K_{\text{ATP}}}) \cdot (1 + \frac{\text{fbp}}{K_{\text{FBP}}}) \cdot (1 + \frac{\text{atp}}{K_{\text{ATP}}} + \frac{\text{atp}}{K_{\text{ADP}}})}$$

Selected item: Model: PFK Kinetic model

<https://fairdomhub.org/investigations/56>

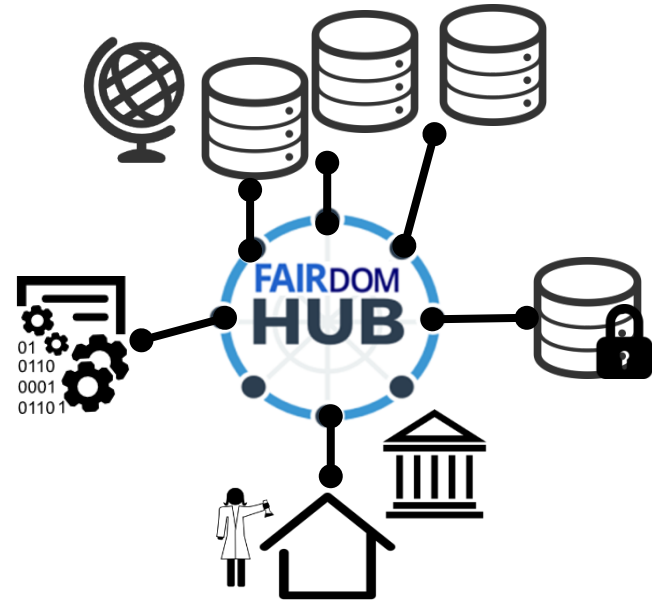
Citation G. Penkler; F. du Toit; W. Adams; M. Rautenbach; D. C. Palm; D. D. van Niekerk; J. L. Snoep; (2014): Glucose metabolism in Plasmodium falciparum trophozoites; FAIRDOMHub.

<http://dx.doi.org/10.15490/seek.1.investigation.56>



Multi-results & Versions  
 Data of many types...Primary,  
 secondary, tertiary...  
 Methods, models, scripts ...  
 Physical objects: samples, strains,  
 specimens

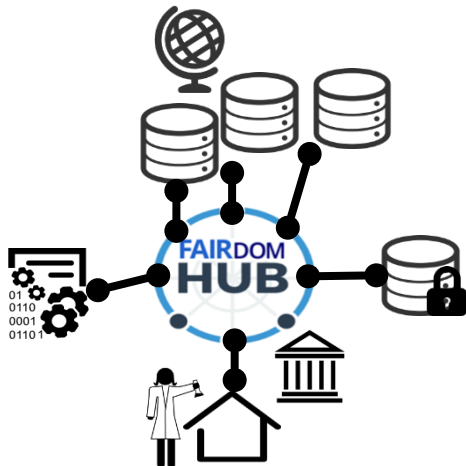
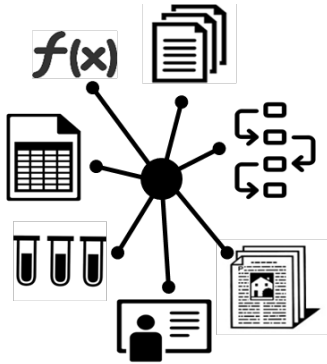
Structured  
 organisation



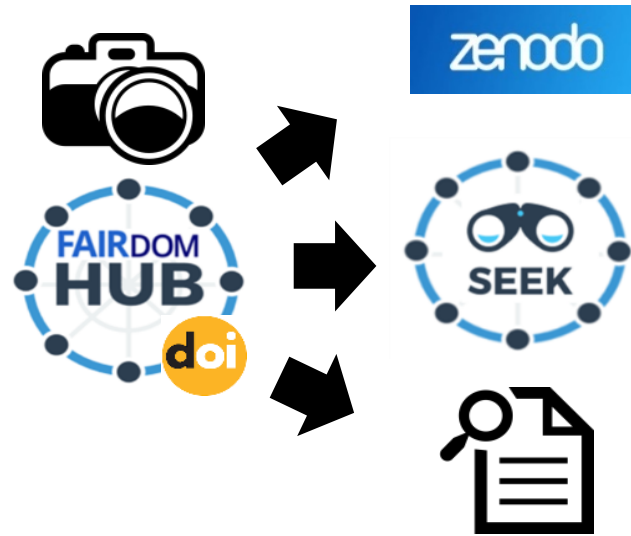
Distributed: Span repository silos,  
 regardless of location and ownership  
 In house and External  
 Multi-site + multi-stewardship

Retaining context  
 over fragmentation

# Spanning across the Ecosystem

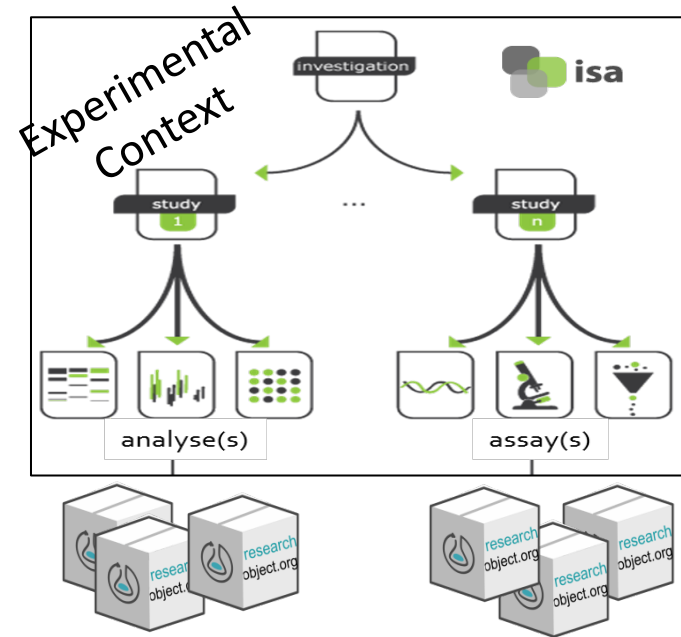
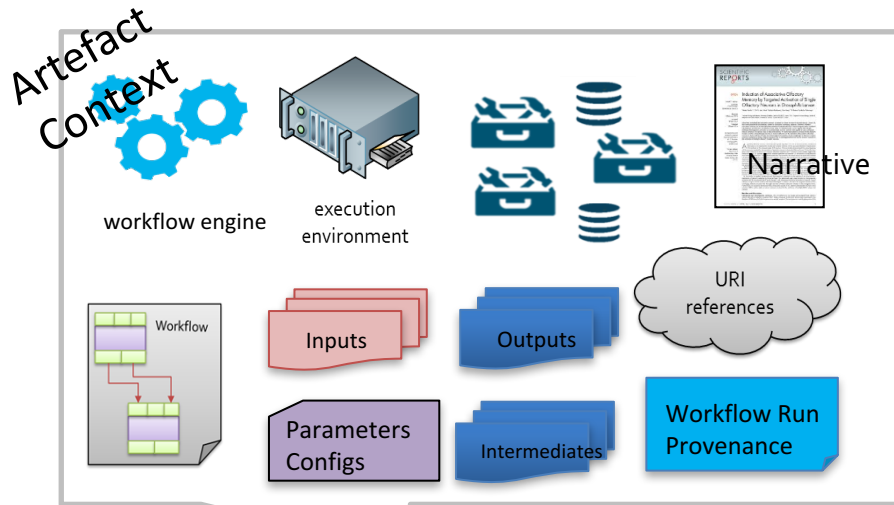


Context

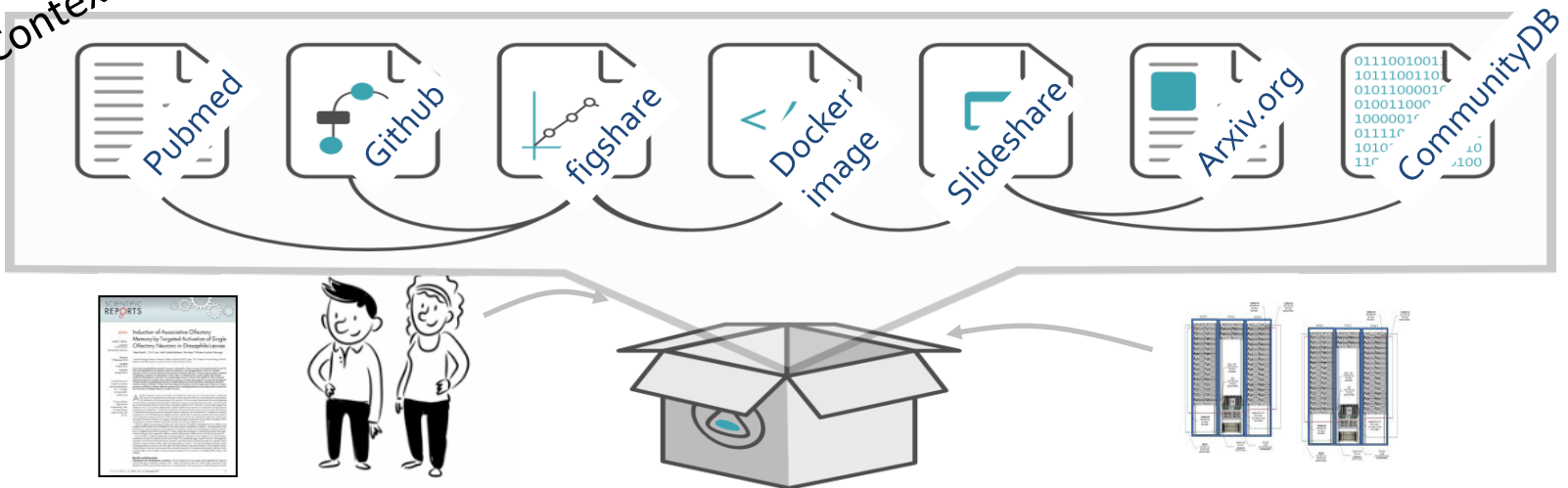


Publishing & Exporting  
Snapshot versions and  
elements  
DOI proliferation

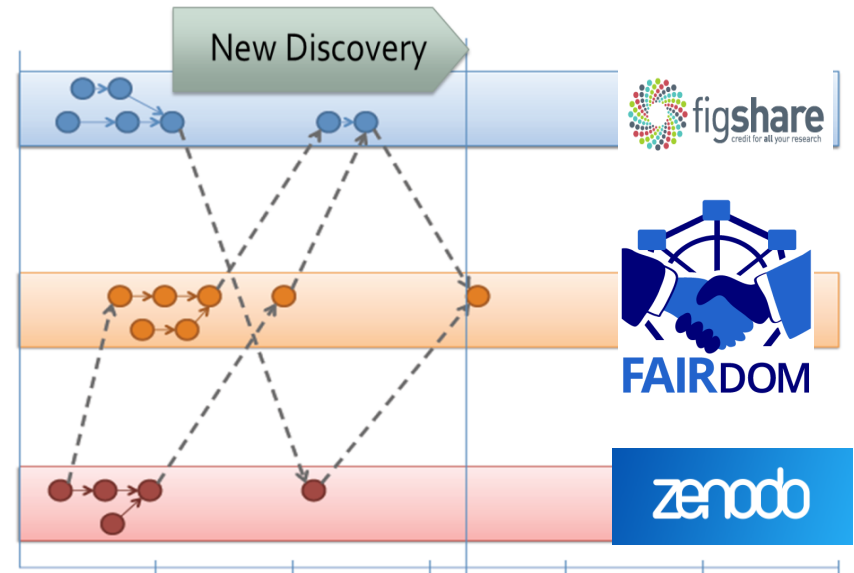
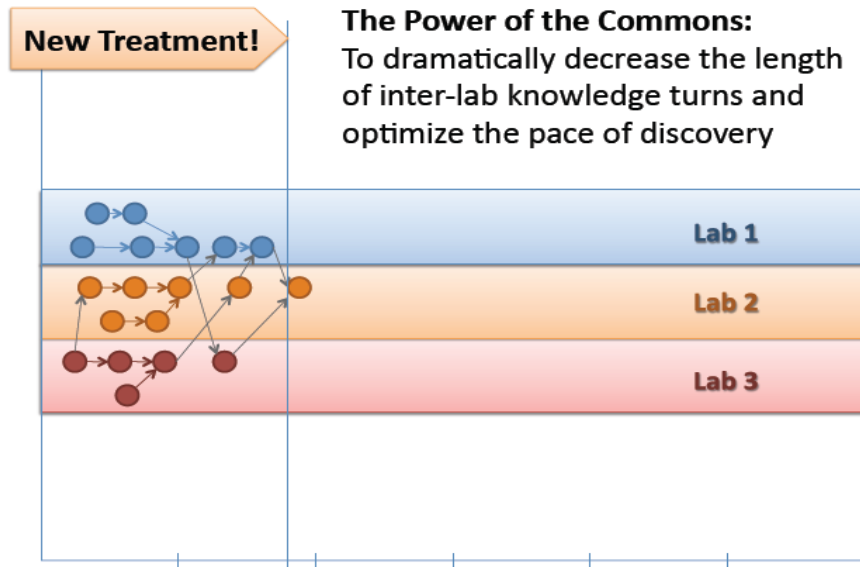
# Digital Object Perspectives

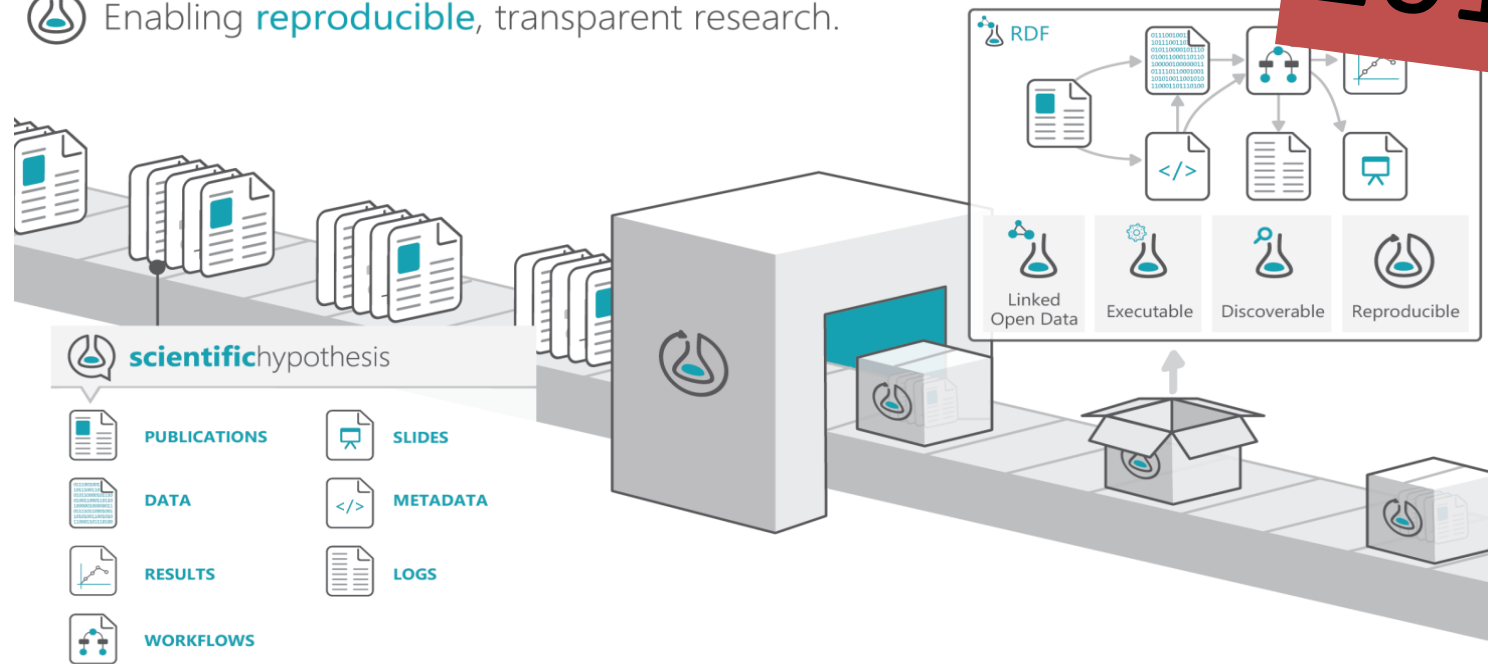


**Repository Context**



preserved, portable research products.  
inter-platform exchange  
multi-platform content & context dependencies





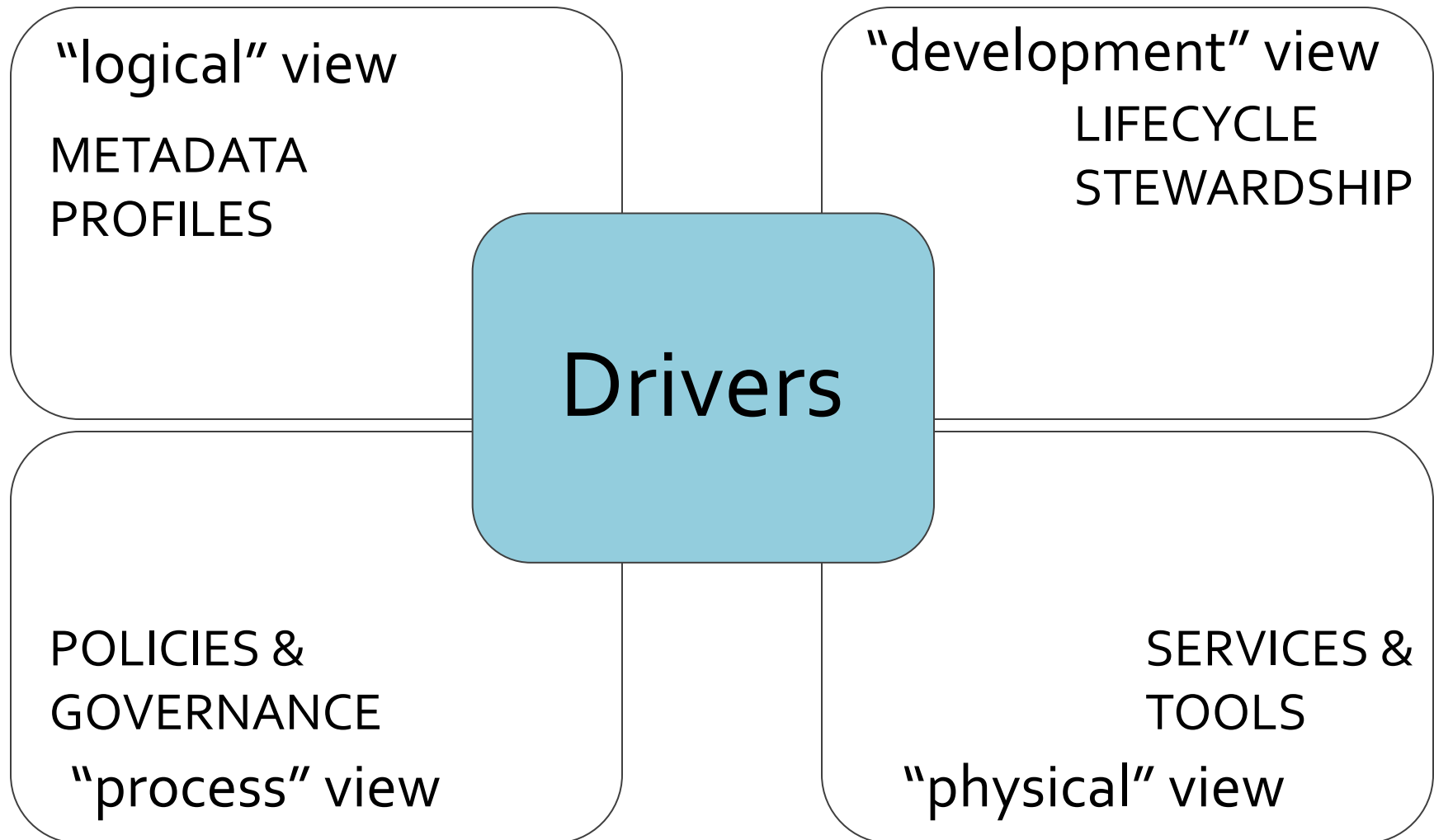
- **Data** used and results produced in experimental study?
- **Methods** employed to produce and analyse that data?
- **Provenance** and settings for the experiments?
- **People** involved in the investigation?
- **Annotations** about these resources, to improve understanding & interpretation

Atomicity,  
Granularity,  
Aggregation  
Composition  
Fragmentation  
Versioning  
Forking  
Cloning  
Portability  
Dependency  
management

Research Objects  
Analogous to  
Software  
artefacts and practices  
rather than  
Data or Articles

# Research Objects

## 4+1 Architecture Framework



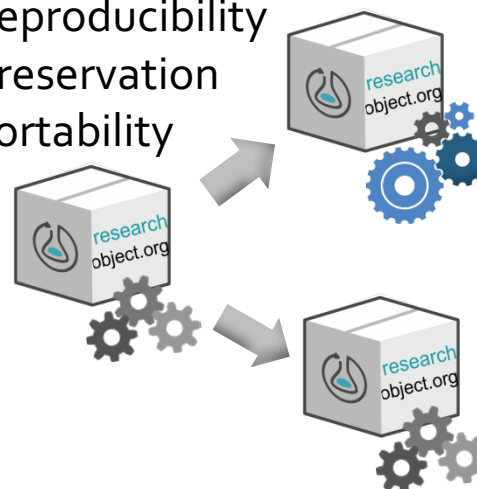
# Drivers within the Ecosystem

Commons & Catalogues  
Publishing,  
Exchange between  
people and platforms  
Sharing,  
Training



Active Research Release  
Evolution & Snapshots  
Remixing, Comparison, Review  
Automated processing

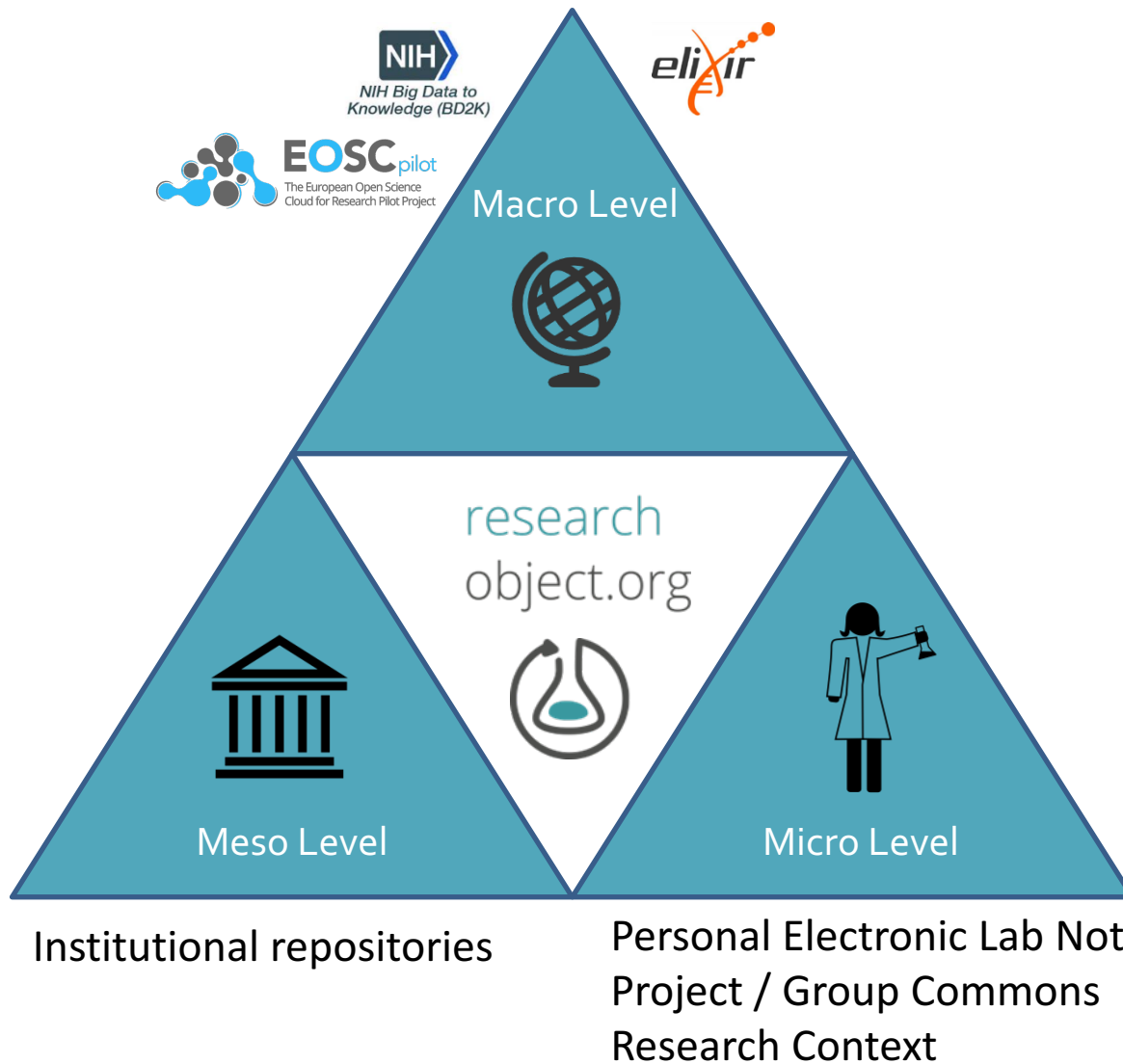
Replication  
Reproducibility  
Preservation  
Portability



Conservation  
Repair  
Archive

(Inter)National Commons  
Public Community Archives  
Publishers

ROs working  
across the  
ecosystem



Boundary  
Objects

# FAIR ROs Desiderata

Use Open Standards.

Incremental.

Machine-processable.

**Graceful degradation.**

Technology Independent.

Multi-platform.

Commodity tooling.

The least possible.

The simplest feasible. Low tech.

Low user overhead and thin client

# Manifests of Metadata

## Manifest Construction

**Identification**  
to locate and cite  
**Aggregates**  
to enumerate  
& link together  
ROs and elements  
**Annotations**  
about RO, elements &  
their relationships



## Manifest Profile Description

**Type Checklists**  
what should be there  
**Provenance**  
where it came from  
**Versioning**  
its evolution  
**Dependencies**  
what else is needed



Container

# Standards & COTS Platforms

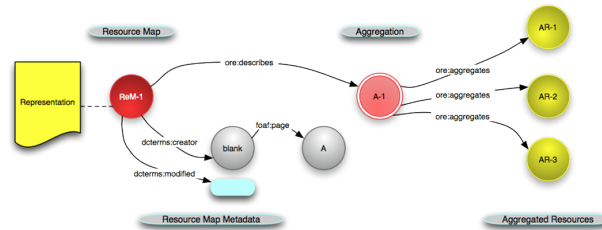
*Identifiers:* URI, RRI, DOI, ORCID



Identifiers.org

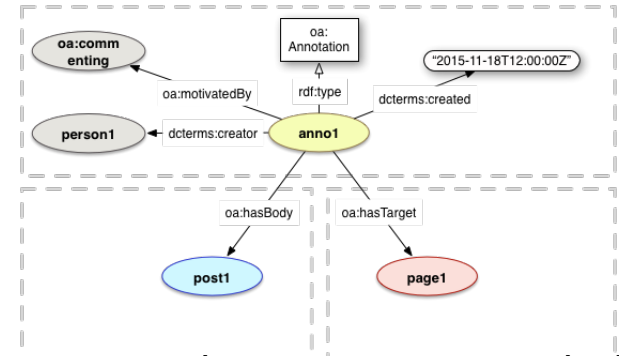


Construction



**Open Archives Initiative**

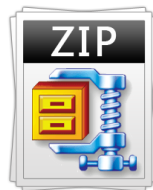
Object Exchange and Reuse  
*Aggregation*



**W3C** Web Annotation Vocabulary  
*Annotation*



Container



**BagIt**



**CONDA**



**BIOCONDA**

# Linking across ROs and into the Linked Open Data Cloud



- Recording & linking together the components of an experiment
- Linking across experiments.
- Linked ROs
- Semantic Web + Digital Objects



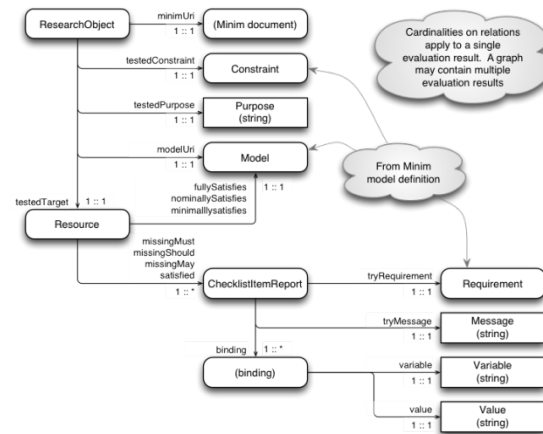
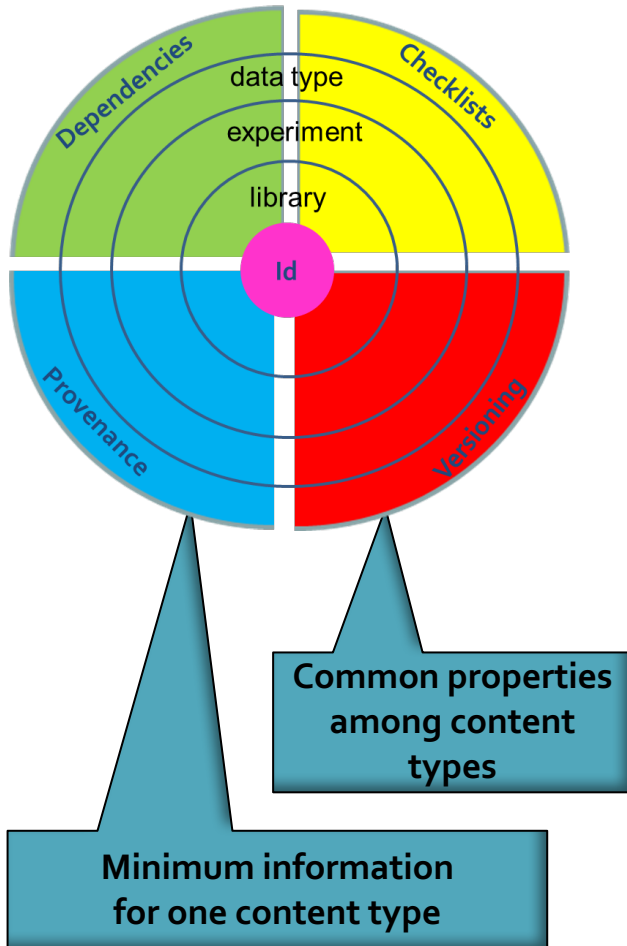
# Goldilocks Profiles & Progression Levels to define and interpret content

its evolution	where it came from	what else is needed	what should be there for types	
<b>Versions</b> 	<b>Provenance</b> 	<b>Dependencies</b> 	<b>Checklists</b> 	Project / Lab Specific
	<b>wfprov</b> <b>ProvONE</b> 			Community-based Types, Context
				All



# Profiles & Progression Levels

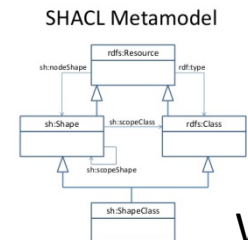
simplify the solution space  
but still encode data types



Minim model for  
defining checklists



<http://purl.org/minim/description>



W3C  
Shape  
Specs

# Validation and Monitoring Tools

## interpret the content

Overview Content Relations Quality Notifications History

Earth Science data-centric RO key requirements

Reload

**✗** Target **Analyzing gene deregulation in Huntington's disease with respect to epigenetic information** *does not satisfy* checklist for *ready-to-release*.

- ✓ Research Object has title
- ✓ Research Object has description
- ✓ Research Object has creator
- ✗ Research Object does not have editor
- ✗ Research Object does not have indicated purpose
- ✓ Research Object has subject
- ✗ Research Object does not have copyright holder
- ✓ Data artifact is present (e.g., dataset, document, file, image)
- ✗ One or more data artifact has no format indicated, including

Earth Science workflow-centric RO key requirements

Reload

**✓** Target **UNAVCO GPS Timeseries** *minimally satisfies* checklist for *ready-to-release*.

- ✓ Research Object has title
- ✓ Research Object has description
- ✓ Research Object has creator
- ✓ Research Object has copyright holder
- ✓ Research Object has keywords
- ✓ Research Object has sketch
- ✓ Research Object has subject
- ✓ Research Object has ancillary documentation
- ✓ Workflow resource is present
- ✗ No workflow definition found. The wf definition is the resource understood by a wf engine (e.g., t2flow, ga resource of type Workflow. To solve this issue in ROHub: select the resource annotated with type Workflow "Annotate and Transform"; otherwise add the annotation (hasWorkflowDefinition).
- ✗ No workflow run has been indicated. To solve this issue, aggregate an execution run provenance in ROH generated RO bundle; otherwise aggregate and annotate (WorkflowRun) the provenance resource
- ✓ All Web services found are accessible (or none found): (no values)
- ✗ No workflow process software dependencies found
- ✓ All processes (e.g., workflows) have input defined
- ✓ All processes (e.g., workflows) have output defined
- ✓ All data artifacts have format indicated
- ✓ All data artifacts have filesize indicated



Workflow-centric RO with provenance requirements

Reload

**✗** Target **Analyzing gene deregulation in Huntington's disease with respect to epigenetic information** *does not satisfy* checklist for *complete*.

- ✓ Experiment hypothesis or research question is indicated
- ✓ Experiment description paper is indicated
- ✓ Experiment design sketch is indicated
- ✓ Workflow is present
- ✗ No workflow definitions found
- ✗ No workflow process software dependencies indicated
- ✗ No workflow service references found
- ✗ No workflow sample run data indicated for any workflow
- ✗ No workflow selected input data indicated for any workflow
- ✗ No workflow results data run indicated for any workflow
- ✓ Experiment conclusions are indicated

See quality history with RO Monitoring Tool



This application has been developed by PSNC as part of the Wf4Ever project  
Research Object Digital Library Portal, version 4.8.16-SNAPSHOT.

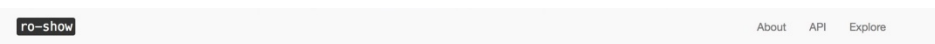
Icons by Glyphicons and the Noun Project.



[Raul Palma]

# Generic Viewing Tooling

## interpret the content



## Research Object Viewer Tool

This is a tool for top level viewing of the resources in a research object and annotations about them. Ro-show is still under development and subject to change without warning!!

### Zip File Input

Choose file No file chosen

Choose a research object zip file to upload

Submit

You successfully uploaded bundle.zip!

<http://localhost:8080/files/bf068bf7e-73ca-457f-8281-516b0d088c9a.zip>  
<http://localhost:8080/files/794db2fa-c126-4c99-9a90-f31409ee3c07.zip>  
<http://localhost:8080/files/4beb2162-1891-456f-987a-4a524d9f00ea.zip>  
<http://localhost:8080/files/f230e94f-c6ba-4a59-8ed8-23371da30323.zip>  
<http://localhost:8080/files/acd651ee-810d-43c9-9b3e-56f598b0b78e.zip>  
<http://localhost:8080/files/4beb2162-1891-456f-987a-4a524d9f00ea.zip>  
<http://localhost:8080/files/f230e94f-c6ba-4a59-8ed8-23371da30323.zip>  
<http://localhost:8080/files/bf068bf7e-73ca-457f-8281-516b0d088c9a.zip>

f230e94f-c6ba-4a59-8ed8-23371da30323

Authors[Jason Walker, Thomas Mooney, Jason Walker, Feiyu Du, Thomas B. Mooney, Susanna Kiwala]

Profile [ Workflow Research Object (can be viewed with Common Workflow Language Viewer)]

Date Created[2017-08-24T10:42:11.587Z]

acd651ee-810d-43c9-9b3e-56f598b0b78e

Authors[Jeremiah H. Savage]

Profile [ Workflow Research Object (can be viewed with Common Workflow Language Viewer)]

Date Created[2017-08-24T10:36:43.975Z]

794db2fa-c126-4c99-9a90-f31409ee3c07

Authors[Pau Andrie]

Profile [ Workflow Research Object (can be viewed with Common Workflow Language Viewer)]

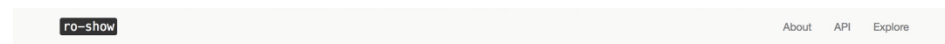
Date Created[2017-11-12T02:07:21.952Z]

4beb2162-1891-456f-987a-4a524d9f00ea

Authors[Thomas B. Mooney, Susanna Kiwala, Jason Walker]

Profile [ Workflow Research Object (can be viewed with Common Workflow Language Viewer)]

Date Created[2017-11-11T01:00:15.225Z]



## preview

authors retrieved from aggregates annotations

Jason Walker  
Thomas Mooney  
Jason Walker  
Feiyu Du  
Thomas B. Mooney  
Susanna Kiwala

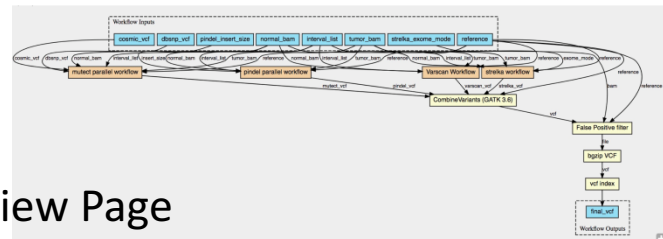
## profile

Workflow Research Object (can be viewed with Common Workflow Language Viewer)

## structure

[under development]

## preview



## Preview Page

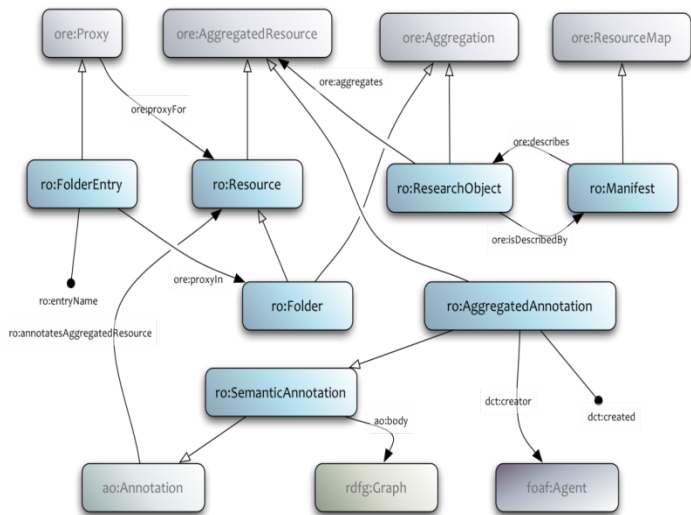
## Interpreting the Content

ro-show: 2020-09-29 ro-show Project and associated Distributed under Apache License, version 2.0 (required attribution notices)

*Making use of these various objects will depend on available infrastructure & tools etc.*

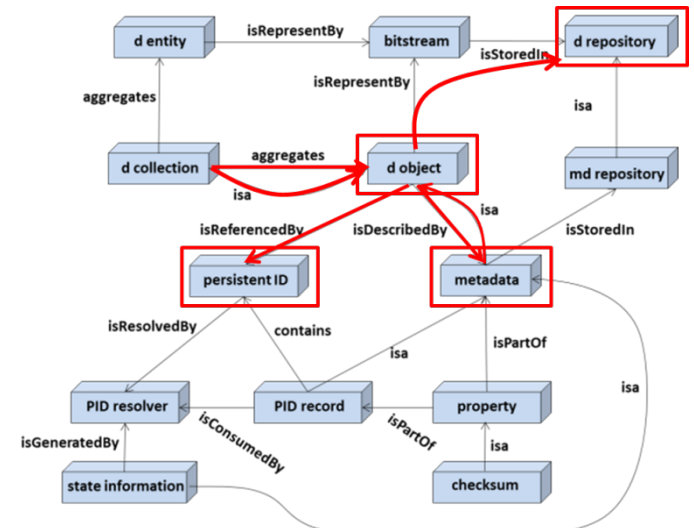
[Lilian Gorea, Oluwatomide Fasugba]

Use the “conformsTo” Property



## Research Object Manifest Model

<http://www.researchobject.org/specifications/>



## RDA Data Foundation and Terminology WG Core model.

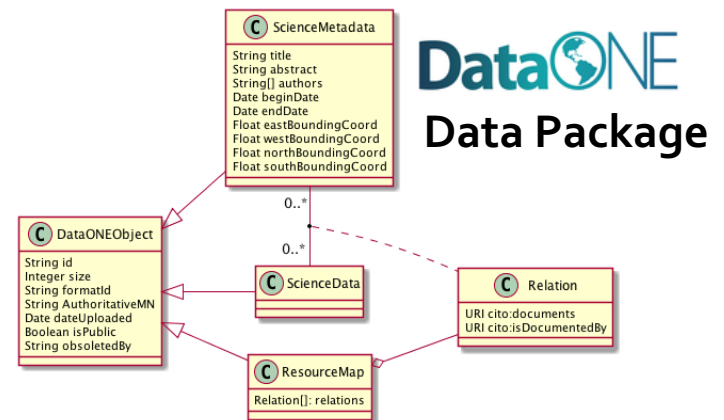


<DATS>

FAIR  
DATA POINT



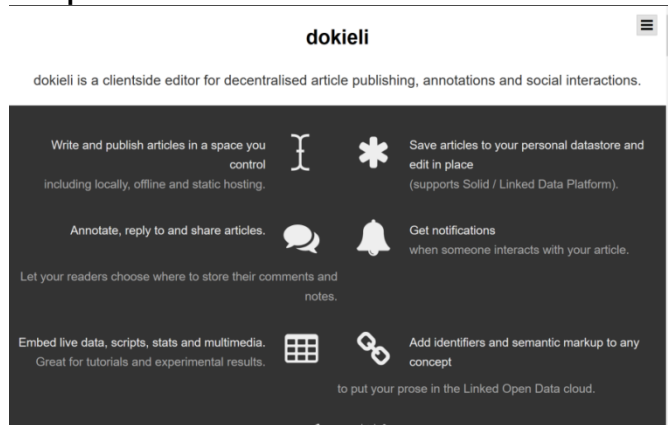
Dockeronto



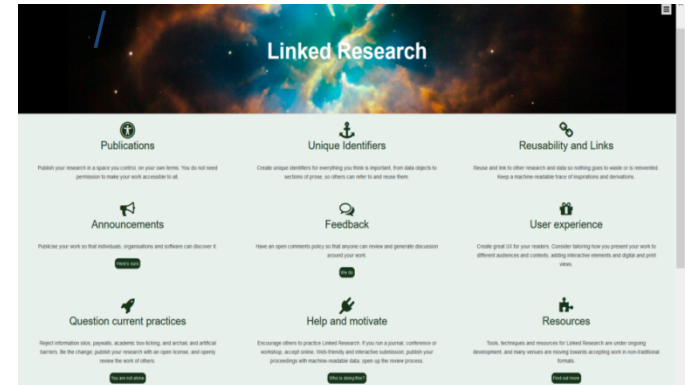
Lots of roads from A to B

# A publishing trend.... JSON(-LD) + schema.org

<https://dokie.li/>



<https://linkedresearch.org>



Manifest: Schema.org,  
JSON-LD, RDF  
Archive: [.tar.gz](#)



Reproducible  
Document Stack project  
eLife, Substance and Stencila

🔗 **DataCrate Specification**

BagIT data profile +  
schema.org JSON-LD  
annotations



# Bioschemas.org

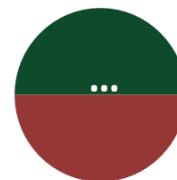
schema.org tailored to the Biosciences

simple **structured metadata markup** on web pages & sitemaps

SCALE!

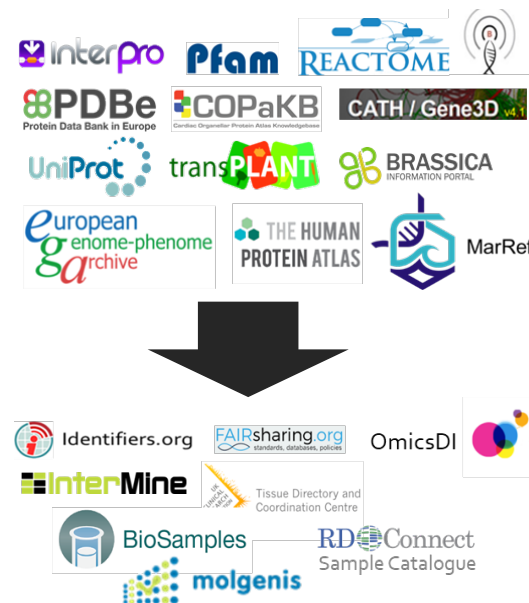
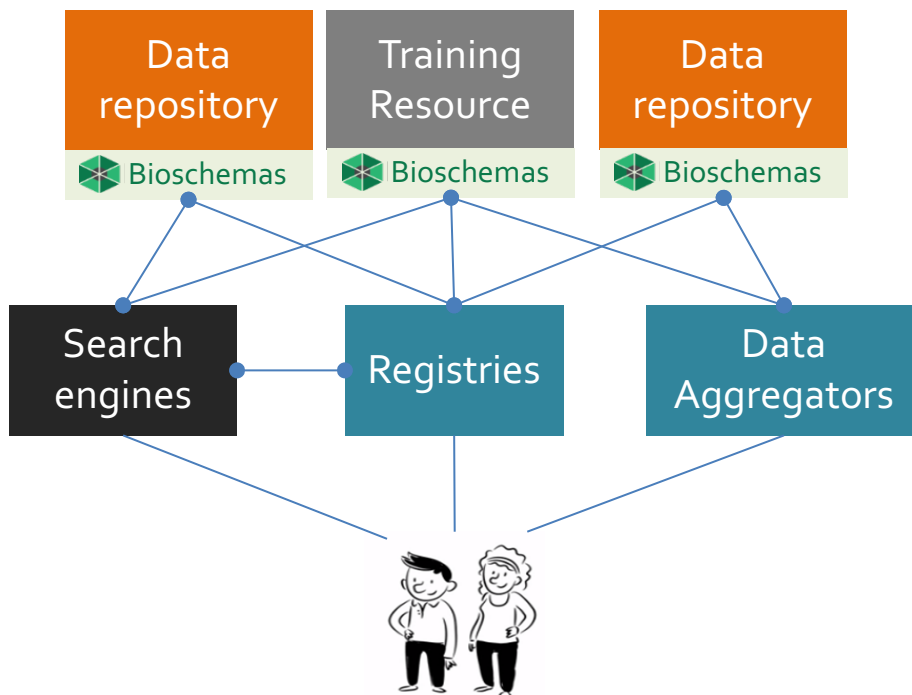
Finding, Citing,  
Metadata Exchange

Bioschemas **Profiles**  
Schema.org **Types**



Standardised  
metadata  
mark-up

Metadata  
published &  
harvested  
without APIs  
or special  
feeds





# Bioschemas.org

schema.org tailored to the Biosciences

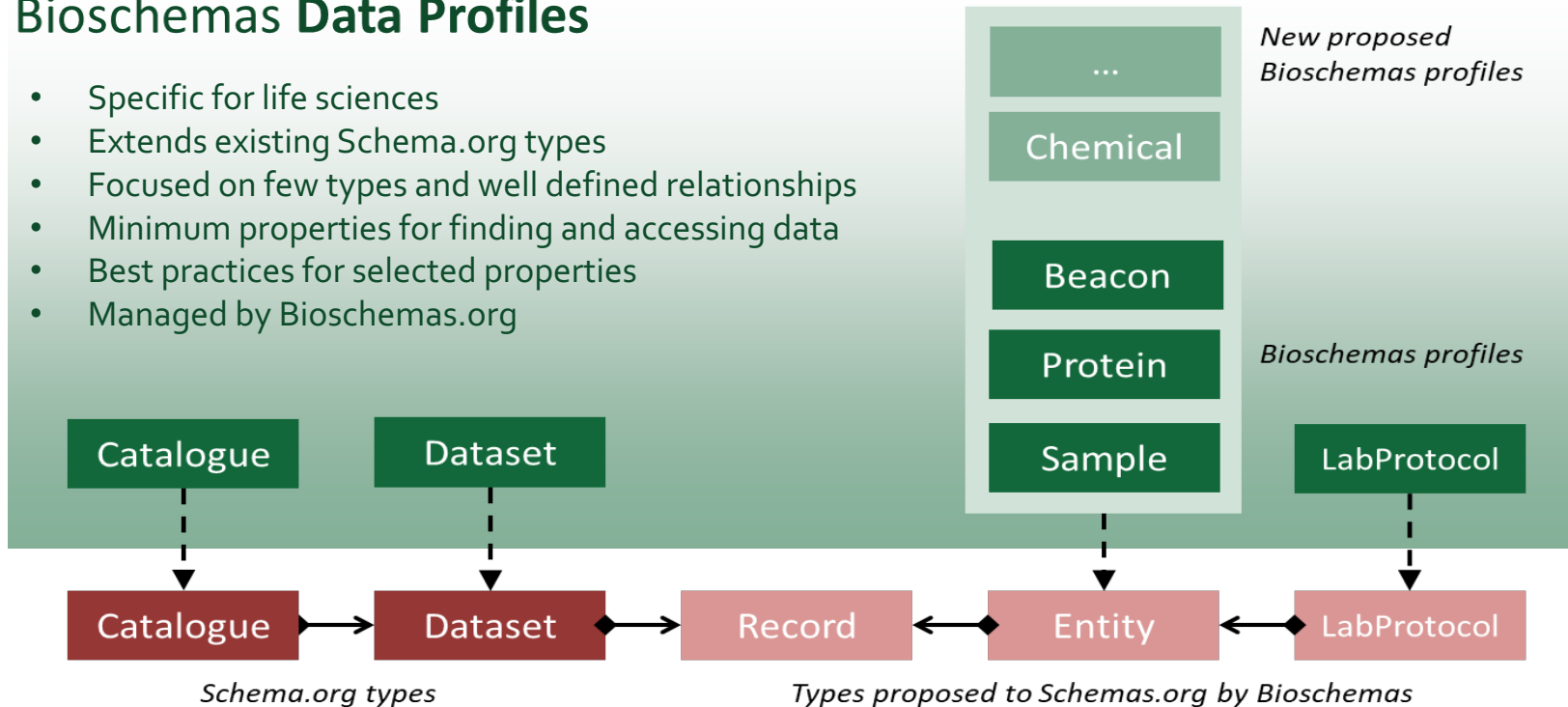
*simple **structured metadata markup** on web pages & sitemaps*

*don't register – harvest & index*

**SCALE!**

## Bioschemas Data Profiles

- Specific for life sciences
- Extends existing Schema.org types
- Focused on few types and well defined relationships
- Minimum properties for finding and accessing data
- Best practices for selected properties
- Managed by Bioschemas.org



## Schema.org Types

- Generic data model
- Generous list of properties to describe data types
- Managed by Schema.org



# Research schemas

# Proposal

Common Research **Types**

Common Research **Profiles**

Specific Research **Profiles**

Bioschemas  
Agroschemas  
Astroschemas  
Earthschemas  
Biodiversityschemas...



**EOSC**pilot

The European Open Science  
Cloud for Research Pilot Project



Bioschemas.org



Tissue Directory and  
Coordination Centre

**Maintain common profiles** across scientific domains focused on finding and accessing data and exchanging metadata in catalogues.  
**Serving Cloud Services & Supporting Boundary Objects**



THE GEORGE  
WASHINGTON  
UNIVERSITY  
WASHINGTON, DC

Precision medicine  
NGS pipelines regulation



Metagenomics  
pipelines and LARGE  
datasets

ISI, USC



Asthma Research e-  
Lab sharing and  
computing statistical  
cohort studies



Public Health Learning Systems



ISA based packaging,  
snapshotting, exporting  
and publishing for Systems  
Biology models



Export RO from WebCAT

U Rostock

Standardised  
packaging of Systems  
Biology models



European Space  
Agency RO Library  
Everest Project

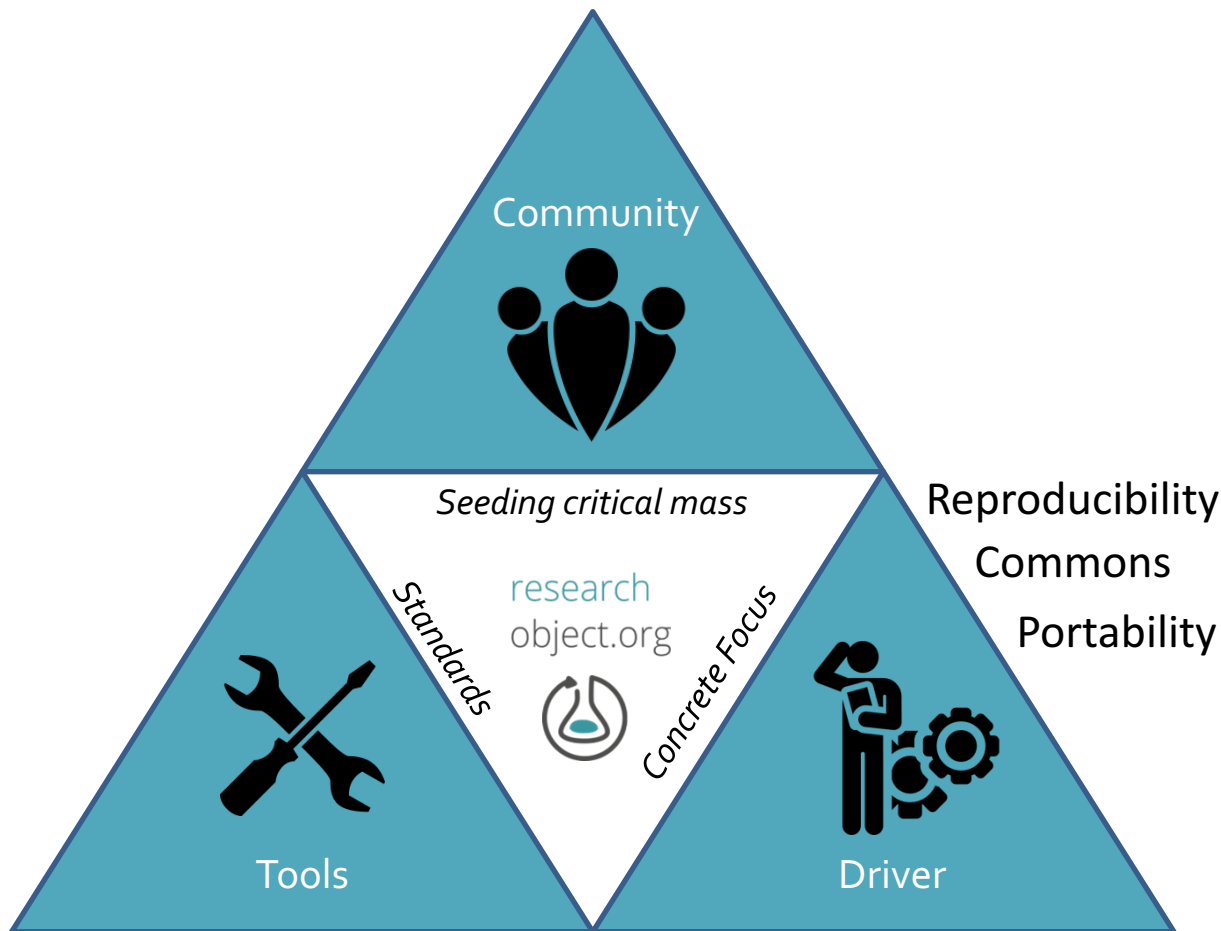


Manifest description of  
CWL workflows

Research Object  
Bundles for Data  
Releases  
Dataset "build"  
tool



Platform & user buy-in from the get-go  
Passionate, dedicated leadership



Easy to make  
Hard to consume  
Generic vs Specific  
Don't be too flexible!

Complex Objects types  
Multi-artefact Objects

# ROs acceptance in the ecosystem

Computational  
Workflows &  
Pipelines

Multi-  
disciplinary  
investigations



[Home](#) / [Publications index](#) / [Fast sampling for quantitative microbial metabolomics: new aspects on cold methanol quenching: metabolite co-precipitation](#)

# Fast sampling for quantitative microbial metabolomics: new aspects on cold methanol quenching: metabolite co-precipitation

[View Publication](#) [Export](#)

## Abstract:

The intra- and extracellular concentrations of 16 metabolites were measured in chemostat ( $D = 0.1 \text{ h}^{-1}$ ) anaerobic cultures of the yeast *Saccharomyces cerevisiae* CEN PK-113-70 growing on minimal medium. Two independent sampling workflows were employed: (i) conventional approach. Metabolites were quantified in different sample fractions (total, extracellular, quenching supernatant, metabolite co-precipitation). The differential method in combination with absolute metabolite quantification by gas-chromatography-mass spectrometry (GC-MS) was used as a benchmark to assess quality of the cold methanol quenching procedure. Quantitative comparison of the two sampling workflows indicates systematic loss of the total mass of various metabolites. The loss of metabolites was correlated with the biomass concentration. The loss of metabolites was also correlated with the biomass concentration. The loss of metabolites was also correlated with the biomass concentration. The loss of metabolites was also correlated with the biomass concentration.

DOI: 10.1007/s11306-014-0700-8

Projects: MOSES

Journal: Metabolomics

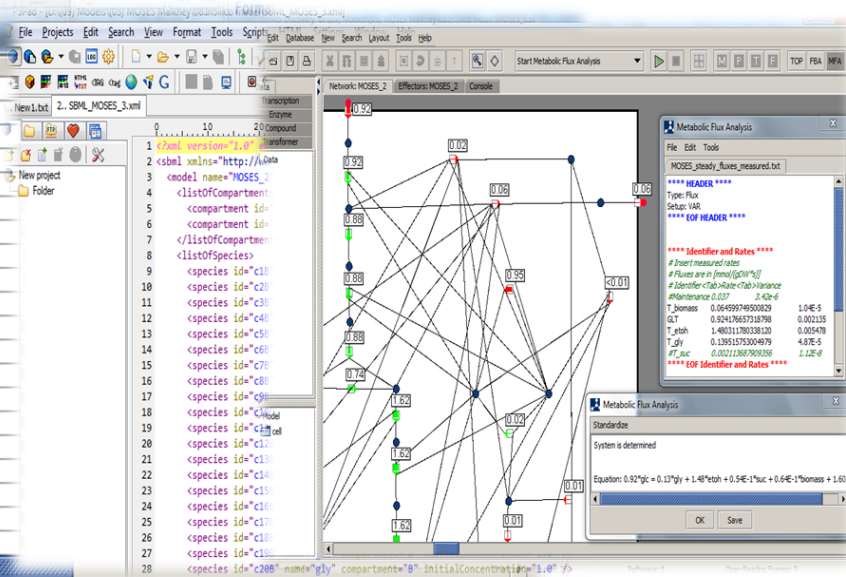
Citation: Metabolomics 11(2) : 286

Date Published: 1st Apr 2015

Authors: Maksim Zakhartsev, Oliver Viehauer, Thomas Horn, Xuelian Yang, Matthias Reuss

Selected: Fast sampling for quantitative microbial metabolomics: new aspects on cold methanol quenching: metabolite co-precipitation  
Type: Publication

J9		ADH		ADK	
A	B				
#	Metabolite				
4	acetic acid	ADH_cell	-1.0	0.0	
5	succinic acid	ADP_cell	0.0	2.0	
6	pyruvic acid	AMP_cell	0.0	-1.0	
7	lactic acid	gap_cell	0.0	0.0	
8	acetaldehyde	dhap_cell	0.0	0.0	
9	CO2	f16p_cell	0.0	0.0	
10	glucose-6-phosphate	IMP_cell	0.0	0.0	
11	fructose-6-phosphate	PRPP_cell	0.0	0.0	
12	dihydroxyacetone phosphate	CO2_cell	0.0	0.0	
13	glycerate 3-phosphate	pep_cell	0.0	0.0	
14	phosphoenolpyruvate	g3p_cell	0.0	0.0	
15	1-glycerol-phosphate	ogp_cell	0.0	0.0	
16		g6p_cell	0.0	0.0	
17		glc_cell	0.0	0.0	
18		gly_cell	0.0	0.0	



Annotation file      Stoichiometric matrix      SBML      Stationary fluxes

## Related items

People (1)   Projects (1)   Investigations (2)   Studies (3)   Assays (3)   Data files (4)   Models (1)

Maksim Zakhartsev

Projects: MOSES, ExtremoPharm, ZuoAI, GenoSyFas, DigiSal  
Institutions: University of Stuttgart, University of Hohenheim, Norwegian University of Life Sciences

Disciplines: Modelling, Experimentalist, Bioinformatics

Roles: Member, Postdoc

Expertise: yeast, fungi, Metabolomics, Proteomics, Stoichiometric modelling, carbon metabolism, Systems Biology, dynamics and control of biological networks, coupling metabolome and environment, rapid sampling experiments, Biochemistry

Tools: parameter estimation, ODE, HPLC, GC and LC-MS analysis of metabolites, Mass spectrometry (LC-MS/MS), continuous cultivation, Enzyme assay, Material balance based modeling, stimulus response experiments, Chromatography, Fermentation, Matlab, Metabolomics, Biochemistry and metabolite analysis



[Maksim Zakhartsev]

# Computational Workflow Research Objects



Community led standard way of expressing **and running workflows** and the command line tools they orchestrate, supporting containers for portability.

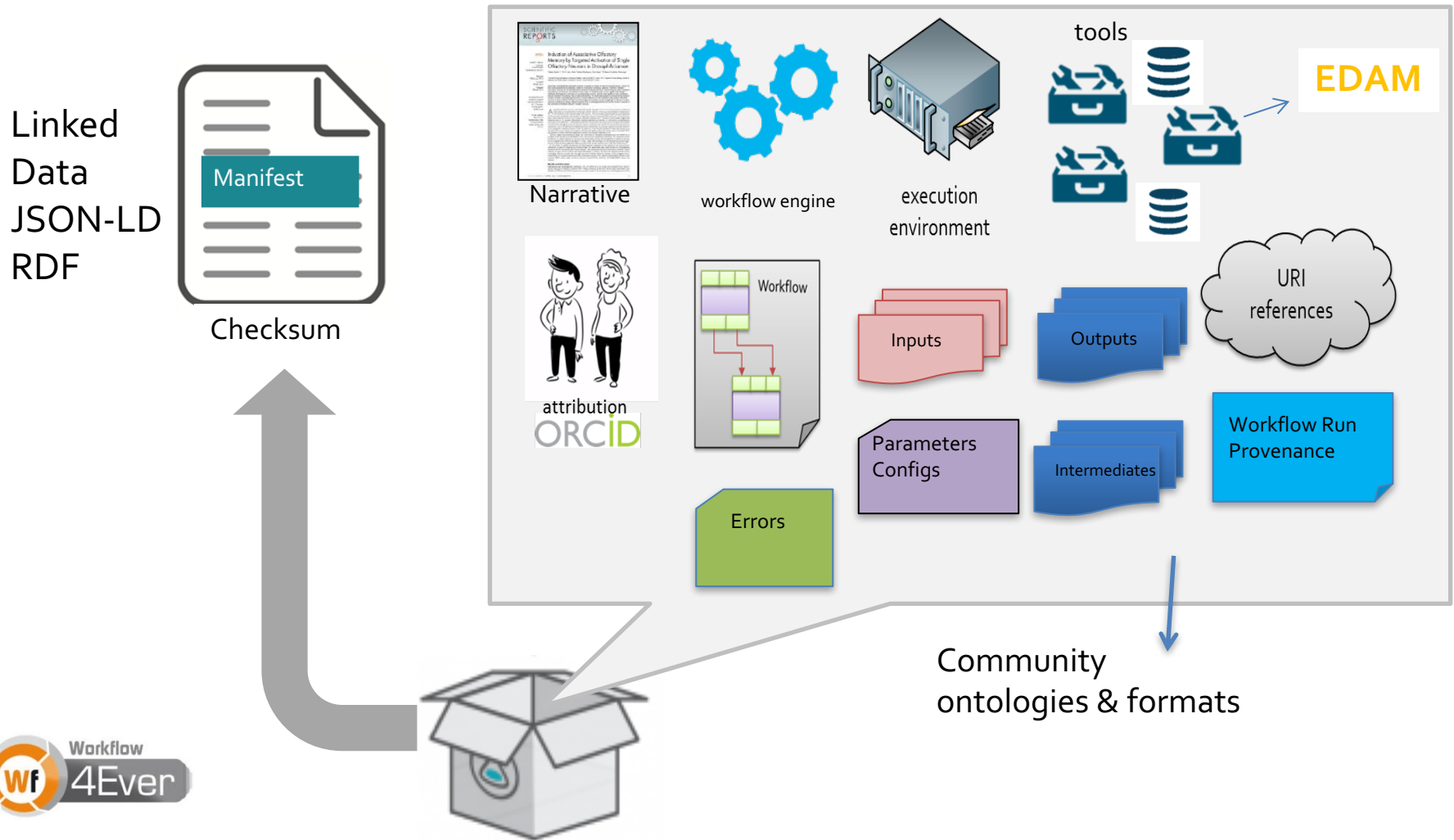


Gathers CWL workflow descriptions together with rich context and provenance using multi-tiered descriptions  
Snapshots the workflow.  
Relates to other objects.



**schema.org**

# Reports + Machines : Workflow Research Objects



Belhajjame et al (2015) Using a suite of ontologies for preserving workflow-centric research objects, J Web Semantics doi:10.1016/j.websem.2015.01.003

Hettne KM, et al (2014), Structuring research methods and data with the research object model: genomics workflows as a case study. J. Biomedical Semantics 5: 41

COMMON WORKFLOW LANGUAGE

About API Explore

### Workflow: bam-genomecov-bigwig.cwl

Fetches 2017-10-17 14:49:08 GMT. Download as Research Object Bundle [?]

creates genome coverage bigWig file from .bam file.

Permalink: [?] <https://oid.org/cwl/view/61c723fe94cda2d1efc7f792979e31423648d59/workflows/bam-genomecov-bigwig.cwl>

Verified with cwltool version 1.0.20

View DOT Download Page

Workflow Inputs: scale, mappedreads, pairchip, strand, split, fragmentsize, input, genomeFile, bigWig, "-bg"

Workflow Outputs: bed\_file, outfile

Tools: genomecov, sort, bigwig

Legend: Inputs/Outputs (blue), Tools (yellow), Nested Workflows (orange), Default Values (purple), Selected (green)

Download as a Research Object Bundle

Permalink to snapshot the GitHub entry and RO identifier

#### Inputs

ID	Type	Label	Doc
genomeFile	File		
input	File		
mappedreads	Double (Optional)		
strand	String (Optional)		
split	Boolean (Optional)		
pairchip	Boolean (Optional)		
bigWig	String (Optional)		
scale	Float (Optional)		
fragmentsize	Integer (Optional)		

Over an **active** GitHub entry for an **actively developing** workflow

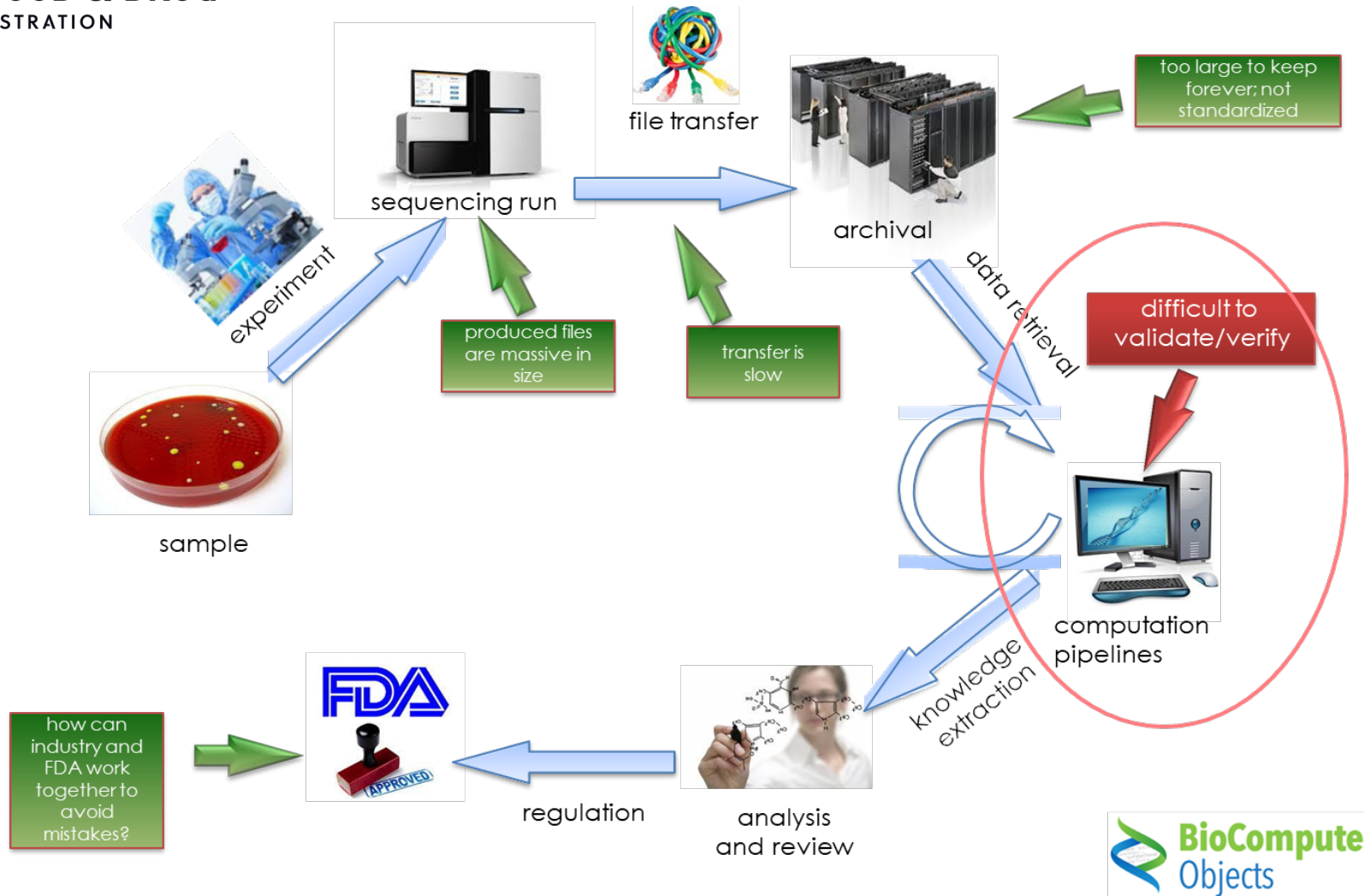
#### Steps

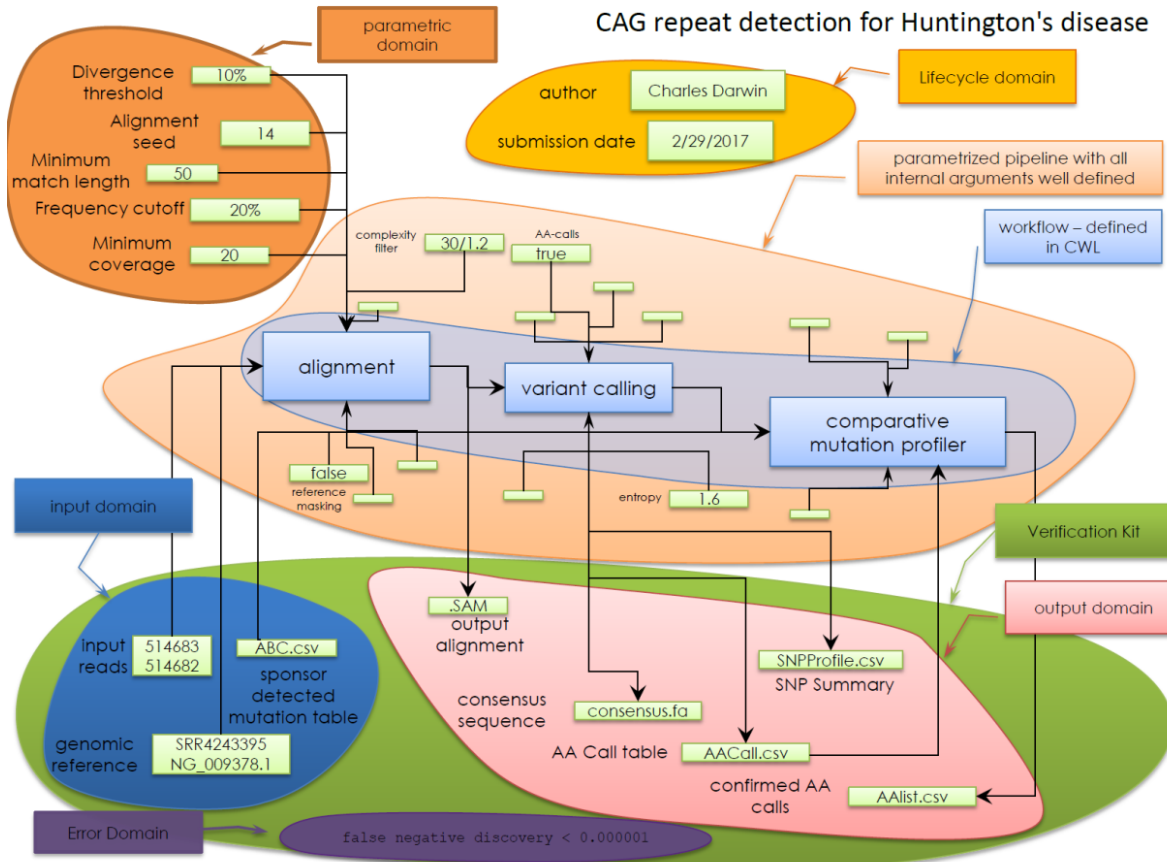
ID	Runs	Label	Doc
genomecov	../tools/bedtools-genomecov.cwl (CommandLineTool)		Tool is used to calculate
bigwig	../tools/ucsc-bedgraphtobigwig.cwl (CommandLineTool)		Tool is used to convert
sort	../tools/linux-sort.cwl (CommandLineTool)		Tool is used to run sort

Special Tooling:  
Common Workflow Language Viewer

# Precision Medicine

## High Throughput Sequencing, from a biological sample to biomedical research and regulation





Linked Data, JSON-LD,  
Ontologies (EDAM, SWO)

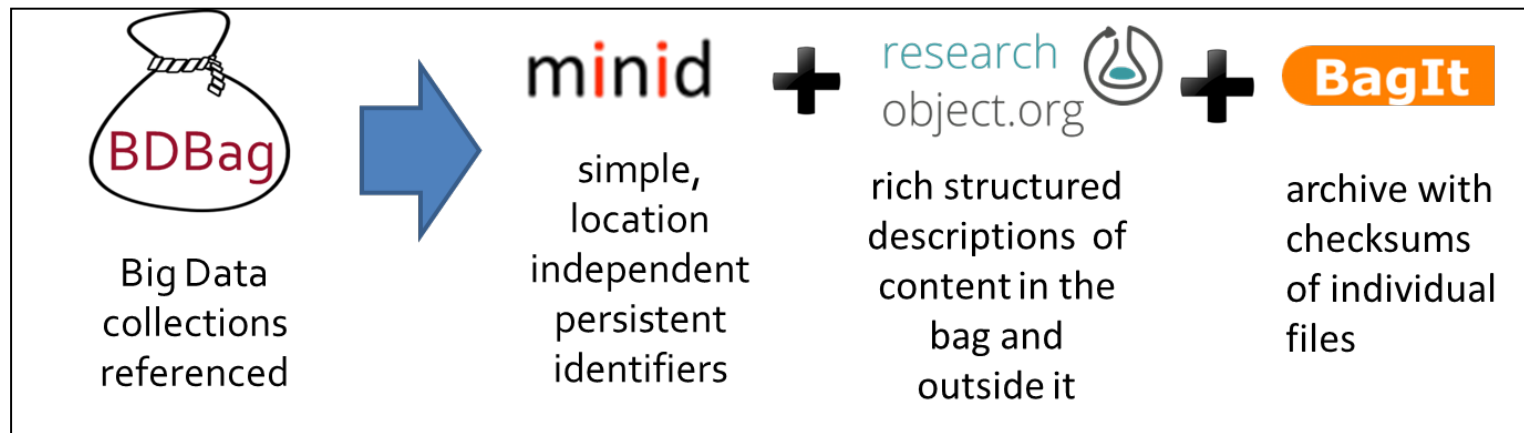


data formats, elements and APIs for EHR & genomics

# Ecosystem of tools and services for big data analysis and sharing in an ecosystem

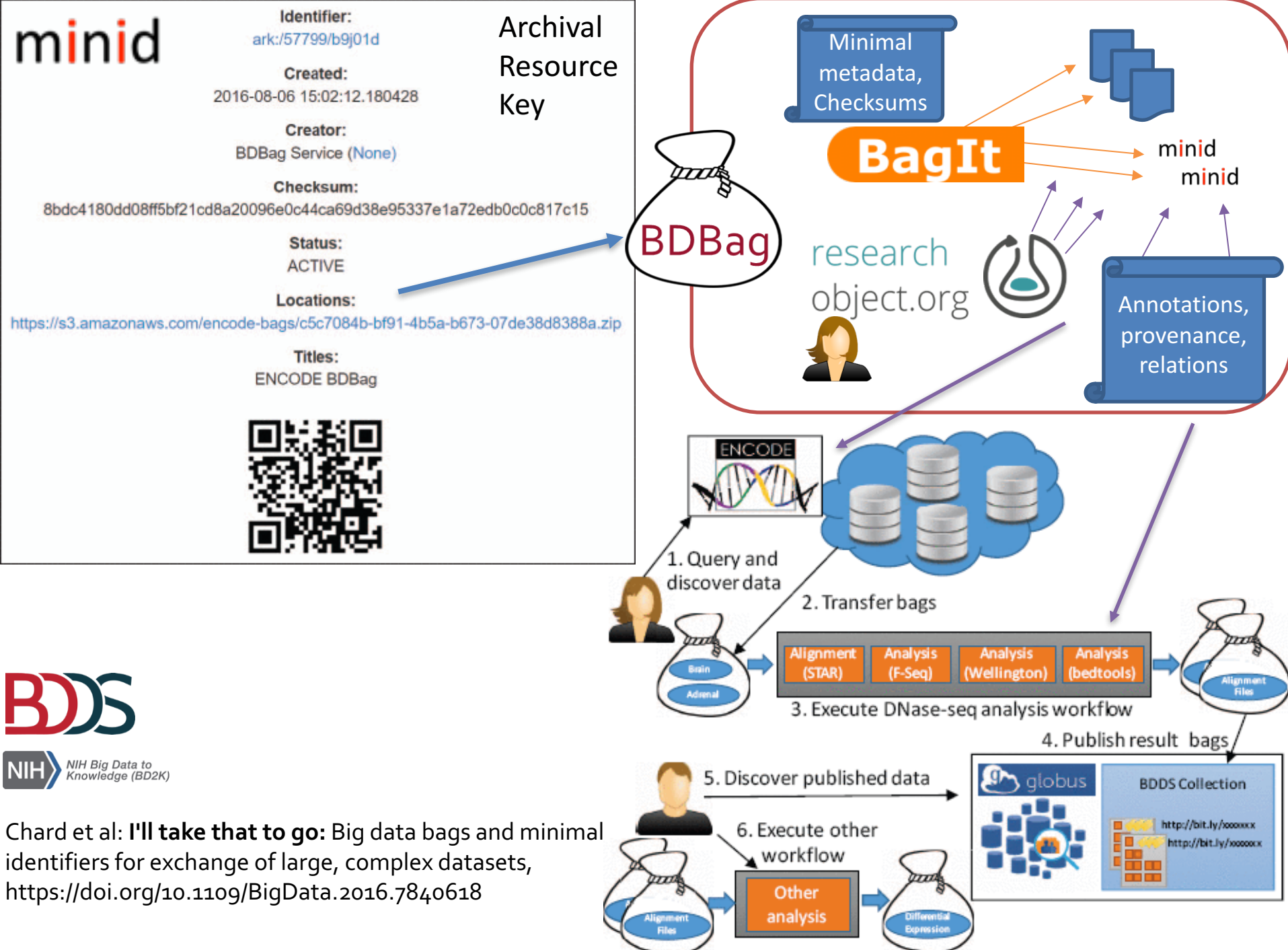
Assemble, share, and analyze large and complex multi-element datasets to integrate into biomedical HTS analytic pipelines

Secure large scale moving of patient data



1000s of images and genome sequences assembled from diverse repositories,

data *distributed* across multiple locations, *referenced* because big and persisted, *efficiently moved* by Grid technologies



# Management in the Ecosystem

whole >  $\sum(\text{independent})\text{parts}$

## Related Items

People (1) Projects (1) Studies (2) Assays (3) **Data files (2)** SOPs (2)


### amino acid auxotrophy for streptococcus pyogenes

SysMO-LAB



amino acid auxotrophy for streptococcus pyogenes in CDM-IMIKRO (not CDM-LAB) as function of final growth yield (OD600 at stationary phase) in 96well plates (200 ul culture)

Contributor: [Araz Zeyniyev](#)

 Download

Created: 5th Apr 2012 at 17:10, Last updated: 12th May 2016 at 16:25

### Characterization of chemostats from S. pyogenes wild type vs. arcA- and glnA-deletion mutants

SysMO-LAB



S. pyogenes was grown in C-limited cultures at pH 6.5 and 7.5 and at a growth rate of 0.05  
The glnA mutant strain shows decreased growth in low glutamine and excess glutamate conditions and no growth at all in low glutamine and low glutamate conditions. Preliminary results of glucose-limited chemostat cultures indicate a reversion of the pH dependency of the shift from homolactic to more mixed acid fermentation:

wild type - lactate/formate ratio at pH 6.5 = 11.8, at pH 7.5 = 2.8


glnA mutant -

...



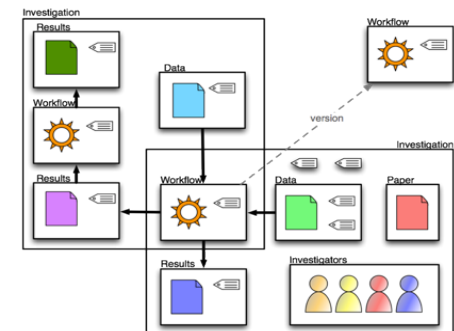
(Show All)

Contributors: [Antje Sieg](#), [Silvio Hering](#), [Martijn Bekker](#)

 Download

Created: 3rd Apr 2012 at 11:26, Last updated: 12th May 2016 at 16:25

Inspirations:  
software development &  
old skool data integration



# Tragedy of the Commons



“The challenge for all the data-commons initiatives — is that many online datasets are annotated with metadata that are simply terrible.... Creating good metadata takes considerable work ....

**When investigators act in their own self-interest, taking short cuts to generate metadata as quickly as possible, we should expect that the overall utility of the resource will decline.**

Manifests of Metadata

- profile making
- template making
- template elements
- auto manufacture
- spreadsheet tooling

<https://metadatacenter.org>

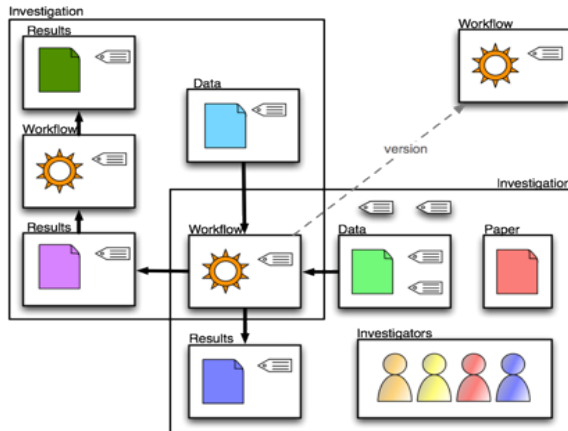


The creation of a data commons requires the ability to deal with extremely varied — and often unanticipated — metadata patterns and data types .... a need for easy-to-use solutions that are generic to provide guidance over the entire life cycle of metadata — streamlining metadata creation, discovery, and access, as well as supporting metadata publication to third-party repositories”

<https://ncip.nci.nih.gov/blog/face-new-tragedy-commons-remedy-better-metadata/>

# Stewardship in a multi-component, evolving ecosystem

Dependencies & Responsibilities with multi-stewardship at different granularities



Who manages the RO and who manages and governs the parts?

Who maintains the manifests?

Delegation and trust!

Expect component rot

A new career?

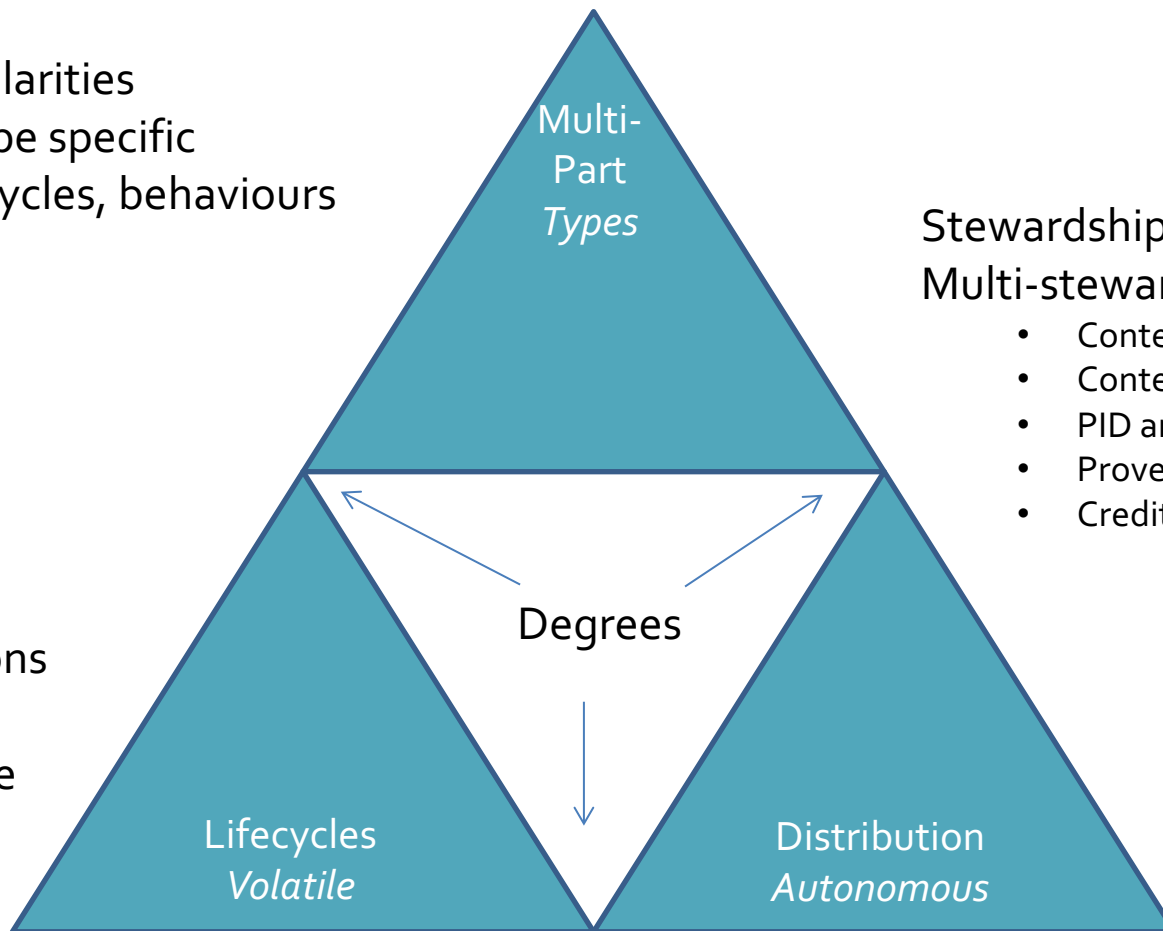


# Multi-Stewardship of ROs and Elements and the stewardship of manifests.....

Different granularities  
Domain and type specific  
standards, lifecycles, behaviours

*Atomicity*  
*Composition*  
*Dependency*

Fixity, verifying  
intended versions  
of contents,  
element change  
detection,  
snapshots

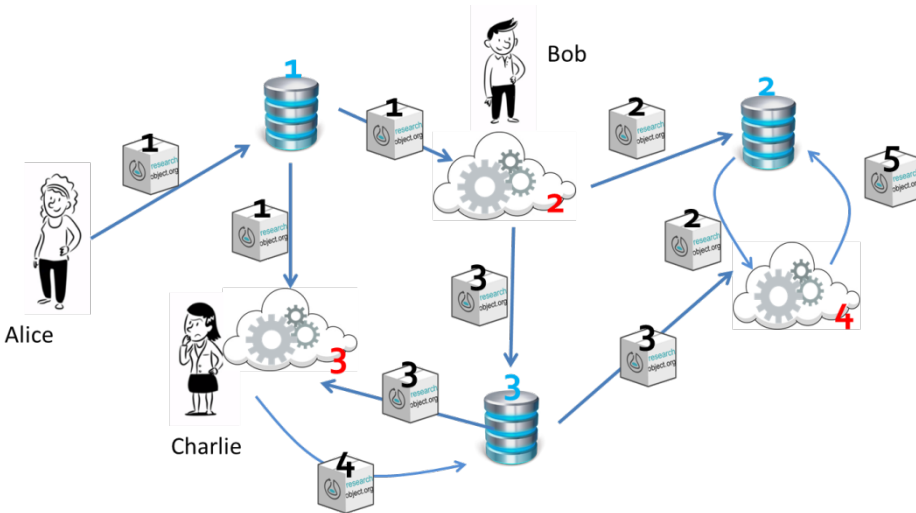


Stewardship hand-offs  
Multi-stewardship guarantees

- Content change
- Content decay
- PID and Resolution services
- Provenance attribution
- Credits

Who is responsible? Spectrum of governance?  
Delegation and degradation.

# Creation, Credit, Curation



Missier, Data Trajectories: tracking reuse of published data for transitive credit attribution, IDCC 2016

## Authenticity, Tamper-proofing

- Hashing & Checksums
- Secure signature & probity services
- Block chain & Ethereum

## DOI proliferation

- Channelling for Counting
- Landing Pages

## Katz and Smith “Contriponents”

- Micro-credit and citation aggregation
- Tracking RO usage & indirect contributions
- Awarding fractional weighted credit to contributors
- Networked Credit maps\*

\*\* D. S. Katz, "Transitive Credit as a Means to Address Social and Technological Concerns Stemming from Citation and Attribution of Digital Products," Journal of Open Research Software, v.2(1): e20, pp. 1-4, 2014. DOI: 10.5334/jors.be

# Trend - bottom up initiatives sheltered by big umbrellas

- Grassroots community activities
- Fostered by Infrastructure Initiatives
- Don't swash the start up!
- Open standards and lightweight
- Practical engineering
- Keeping it simple and real
- **Ramps** rather than Revolution



# Thank You!

Barend Mons  
Sean Bechhofer  
Matthew Gamble  
Raul Palma  
Jun Zhao  
Josh Sommer  
Matthias Obst  
Jacky Snoep  
David Gavaghan  
Stuart Owen  
Finn Bacall  
Paolo Missier  
Phil Crouch  
Oscar Corcho  
Dan Katz  
Arfon Smith

Mark Robinson  
Alan Williams  
Jo McEntyre  
Norman Morrison  
Stian Soiland-Reyes  
Paul Groth  
Tim Clark  
Alejandra Gonzalez-Beltran  
Philippe Rocca-Serra  
Ian Cottam  
Susanna Sansone  
Kristian Garza  
Daniel Garijo  
Catarina Martins  
Alasdair Gray  
Rafael Jimenez  
Iain Buchan  
Caroline Jay  
Michael Crusoe  
Katy Wolstencroft

<http://www.researchobject.org>  
<http://wf4ever.org>  
<http://www.fair-dom.org>  
<http://seek4science.org>  
<http://rightfield.org.uk>  
<http://www.bioschemas.org>  
<http://www.commonwl.org>  
<http://www.bioexcel.eu>

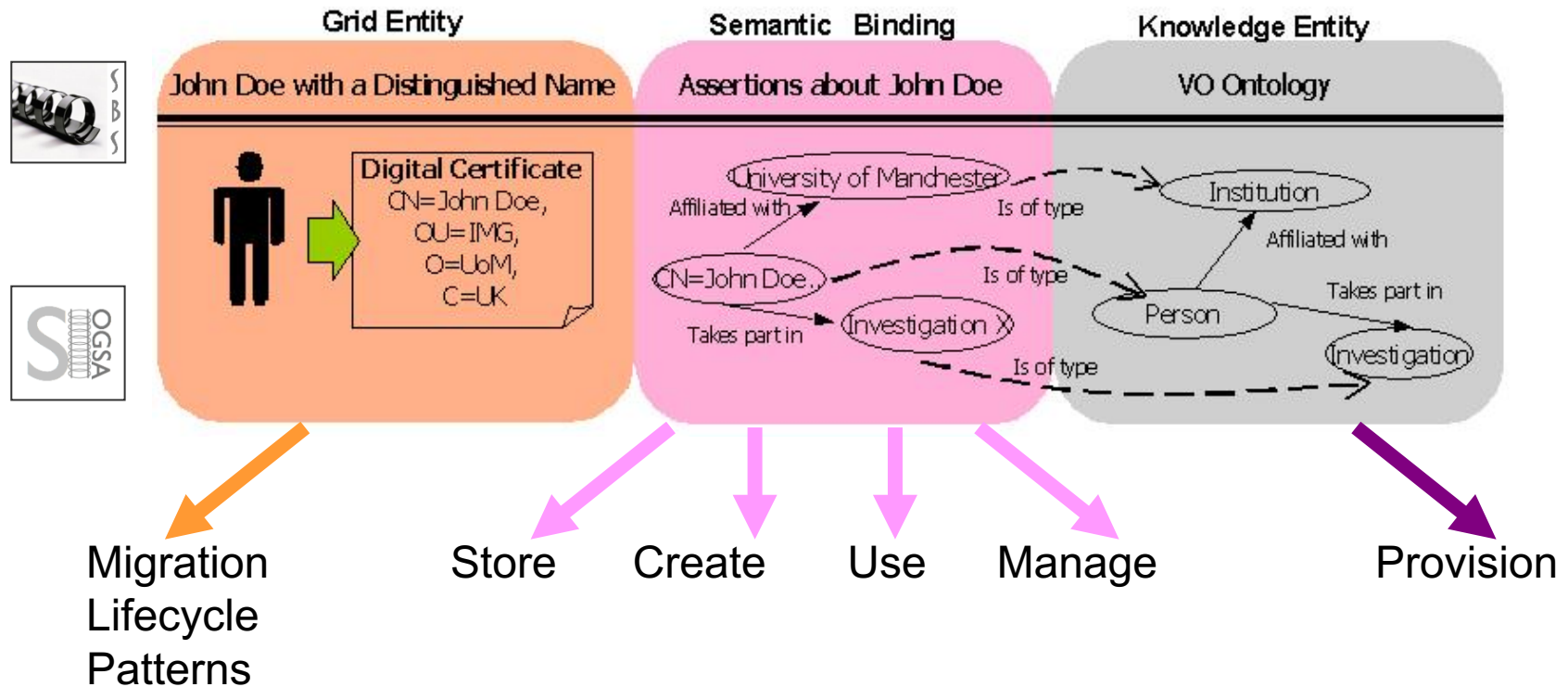
All the members of the Wf4Ever team  
Colleagues in Manchester's  
Information Management Group,  
ELIXIR-UK, Bioschemas



researchobject.org

# Semantic Bindings for non-embedded metadata

## Bind Grid Entities and Knowledge Entities



### An overview of S-OGSA: A Reference Semantic Grid Architecture

Oscar Corcho, Pinar Alper, Ioannis Kotsiopoulos, Paolo Missier, Sean Bechhofer, Carole Goble (2006) <https://doi.org/10.1016/j.websem.2006.03.001>

# Beware the Bucket



**Kathleen Fitzpatrick** ✓

@kfitz

Follow



One of the mistakes that's been made repeatedly in open scholarly communication projects has been the attempt to create the bucket of everything.

5:20 AM - 10 Nov 2017

4 Retweets 34 Likes



5



4



34

<https://twitter.com/kfitz/status/928975762081898496>

# RO based ecosystems

Beyond islands with the odd special ferry-services (e.g. GitHub software releases archived in Zenodo to get DOI)

Bridges to connect anything-to-anything.  
ROs carried between or collected in depots

## Ecosystem evolution

- *Mitigate*
- *Adapt*
- *Migrate*
- *Die*

## Ecosystem incentivisation

- *Love*
- *Money*
- *Fame*
- *By Side effect*