



# Provenance and Linked Data in Biological Data Webs

Jun Zhao

Image Bioinformatics Research Group

Department of Zoology

University of Oxford





# Background

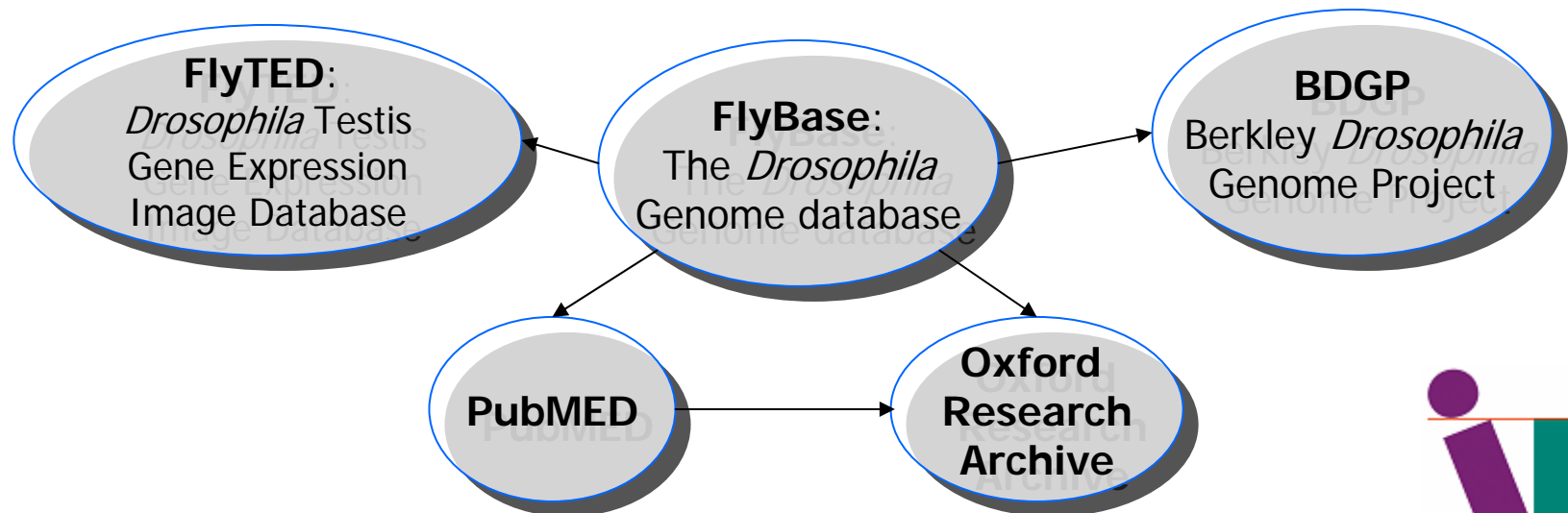
---

- The Image Bioinformatics Research Group
  - User-driven R&D
  - Data integration
  - Data Webs: “use the Web as the native platform in order to enable integrated accesses to datasets including images relating to particular subjects” [*David Shotton. World Wide Science: Promises, Threats and Realities. Data webs for image repositories. Oxford University Press.*]



# FlyWeb: A data web of *Drosophila* data resources

- Link together a number of heterogeneous data resources concerning fruit flies, including gene expression **images** from adult testis (FlyTED) and from embryos (BDGP)
- Initial user studies
- Details: [http://imageweb.zoo.ox.ac.uk/wiki/index.php/FlyWeb\\_project](http://imageweb.zoo.ox.ac.uk/wiki/index.php/FlyWeb_project)





Gene Name: Adh

## FlyTED Testis images of gene Adh

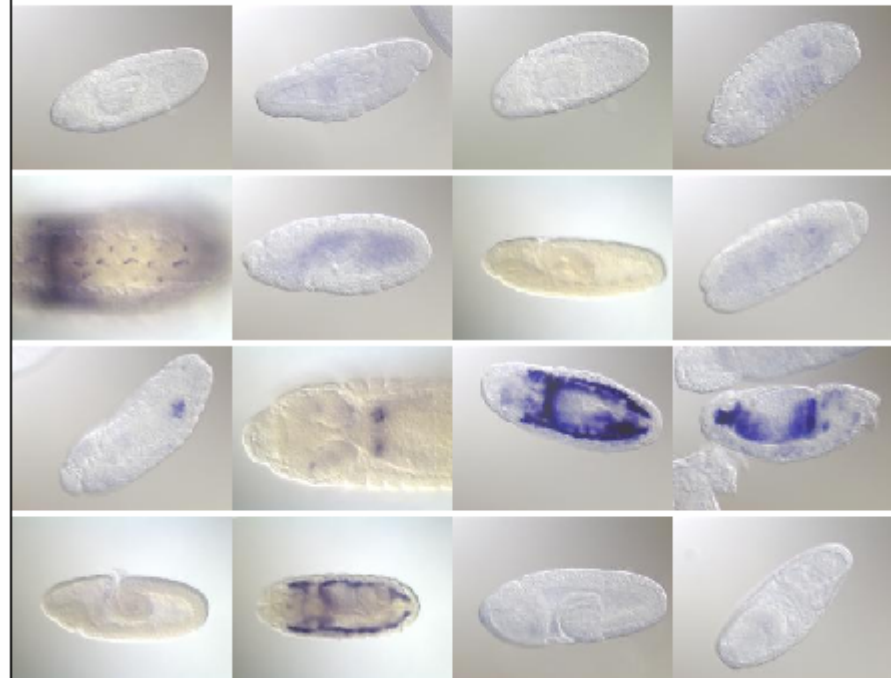


## BDGP Embryonic images of gene CG32954

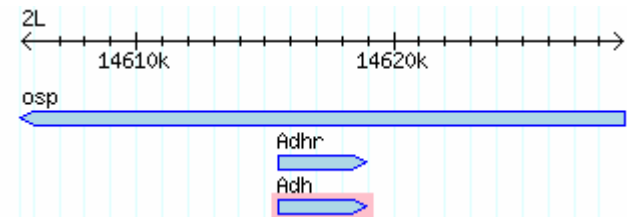
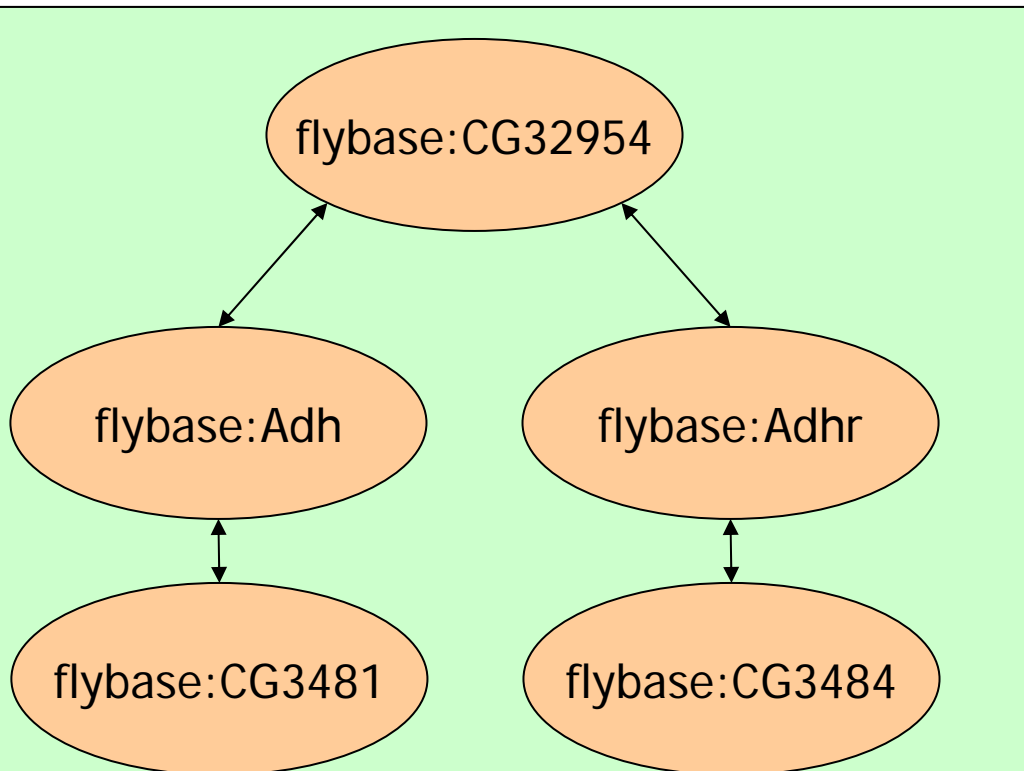
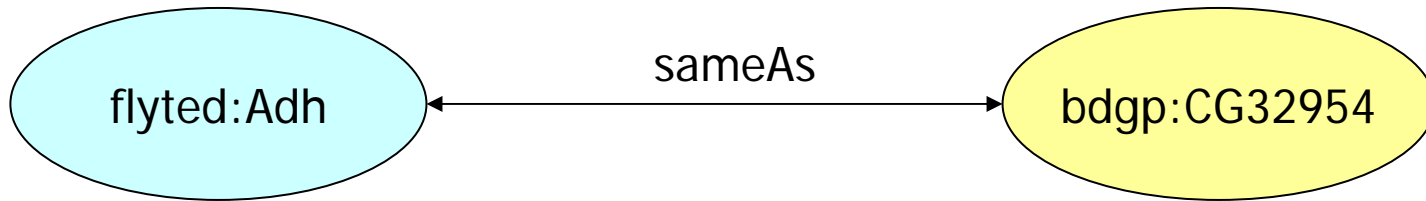


Image results from BDGP.

Found 16 results for query "http://www.fruitfly.org/CG32954".



# Trust on FlyWeb



In FlyBase release 3.2

Since FlyBase release 4.3

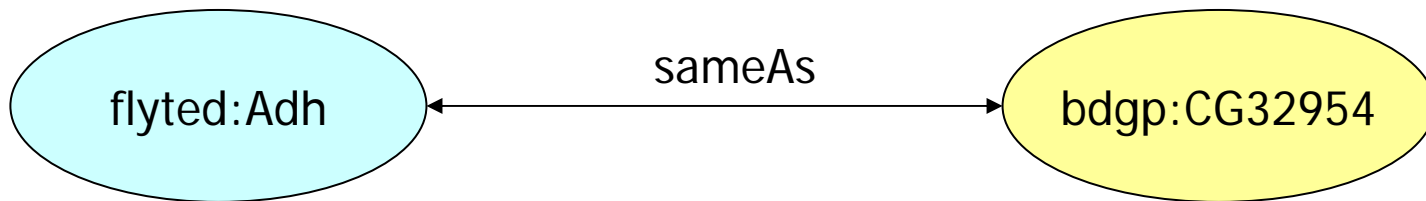
Reference: <http://www.flybase.org>





# Trust on FlyWeb

---



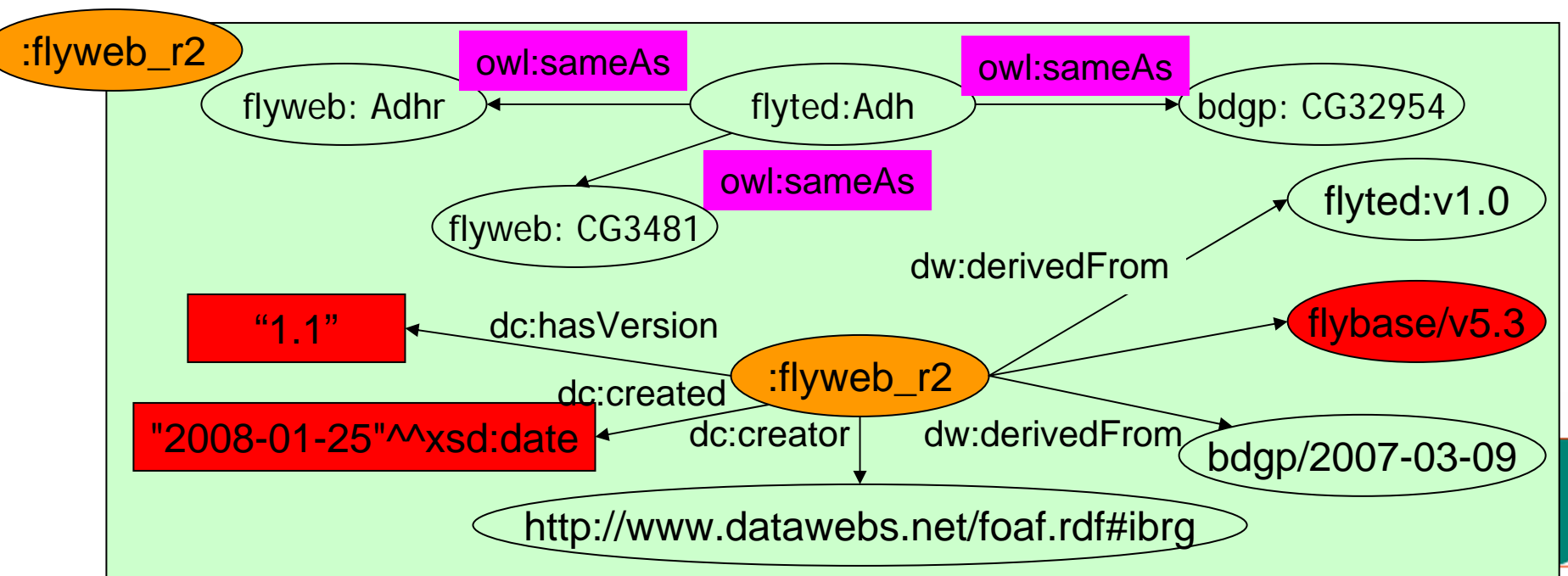
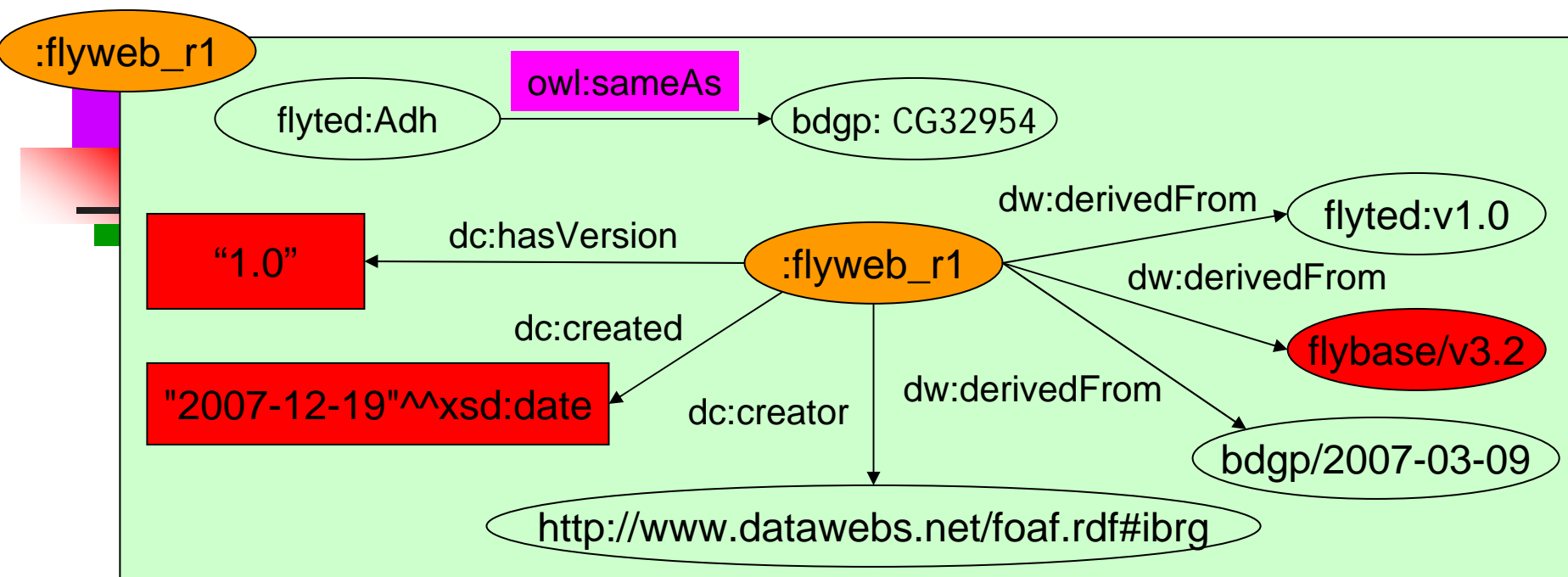
- How the link was built?
- Why these two gene names are the same?
- When this link was created, by whom, using which version of which database;
- What previous links between data items became obsolete, and why.
- How about alternative names for this gene, such as "CG3481", "Dreg-1", etc?



# Provenance for Data Webs

---

- For each release of FlyWeb
  - When it was released
  - Based upon which version of which public database
  - Which data items are links to which other data items





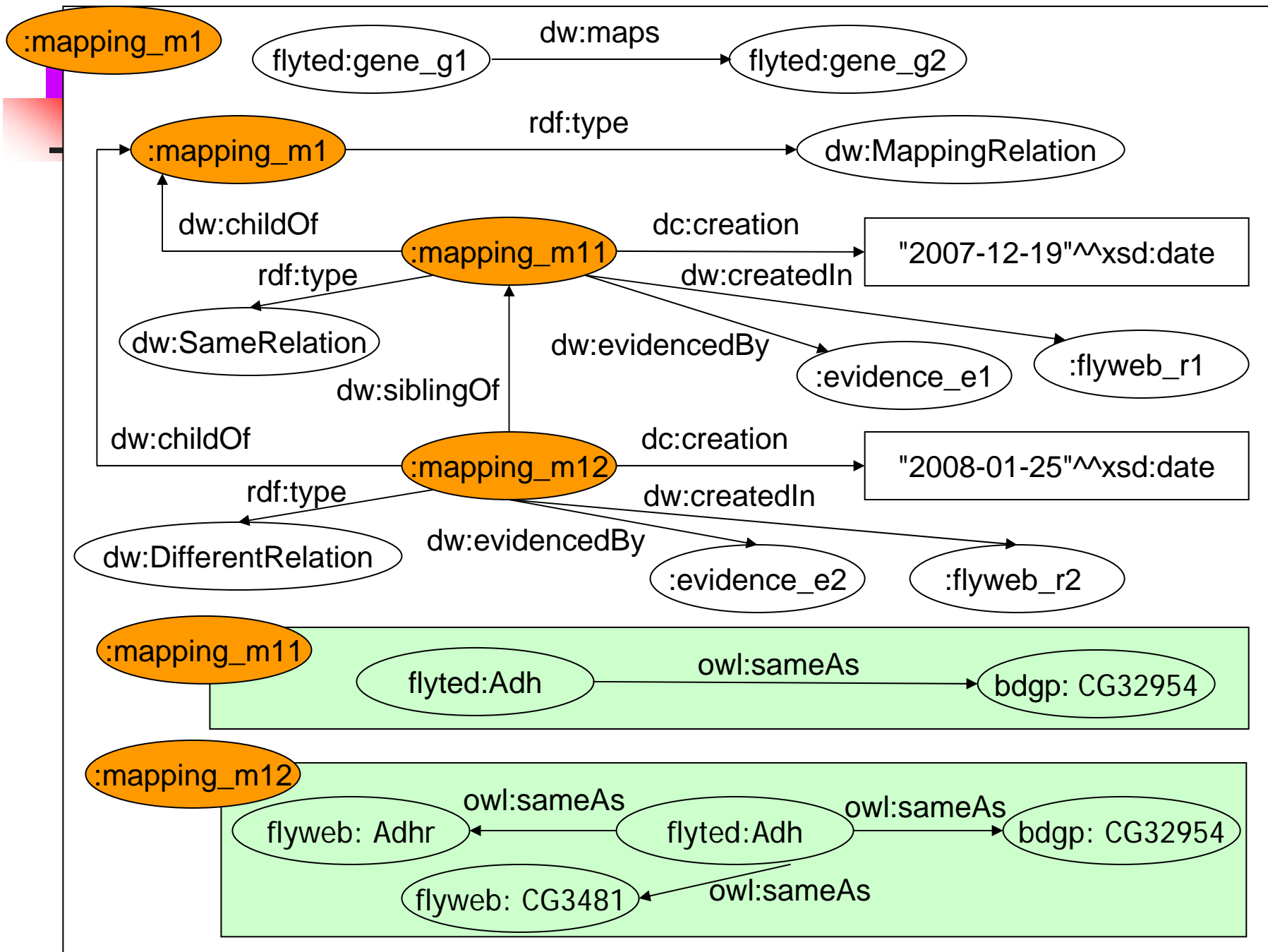


# Provenance for Data Webs

---

- For each pair of linked data items
  - The evidence of the link
  - When the link was built, released, and by whom
  - Which previous links have been created between this pair of related data







---

# Short Demo



## FlyWeb Applications // A // Search FlyTED and BDGP Images By Gene

Adh Go

Image results from FlyTED.

Found 4 results for query "Adh".

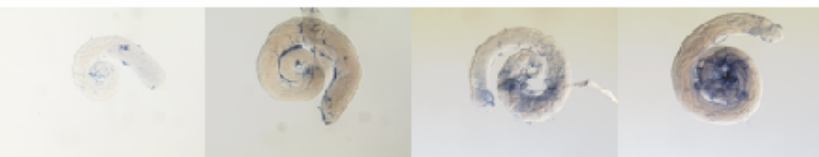
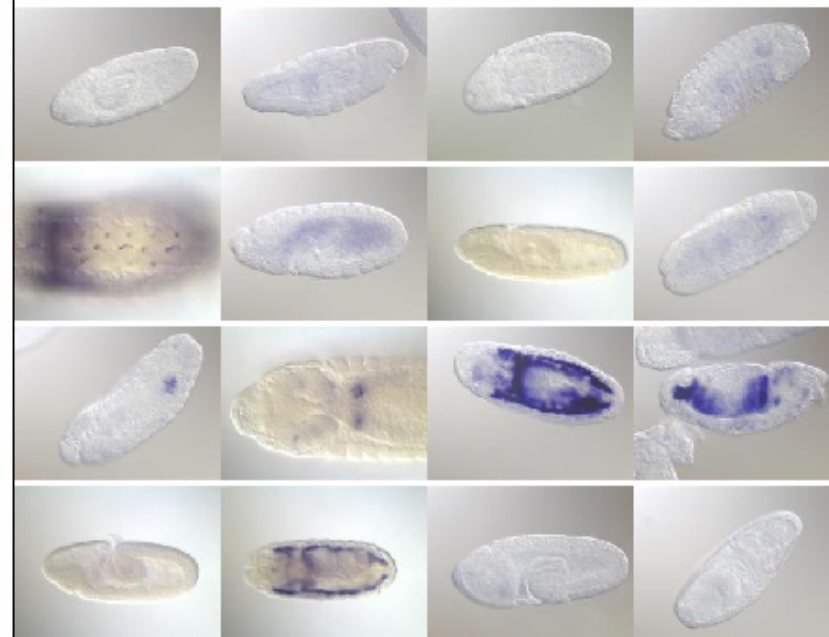


Image results from BDGP.

Found 16 results for query "http://www.fruitfly.org/CG32954".



## FlyWeb Applications // A // Search FlyTED and BDGP Images By Gene

CG3481

Image results from FlyTED.

Found 4 results for query "Adh".

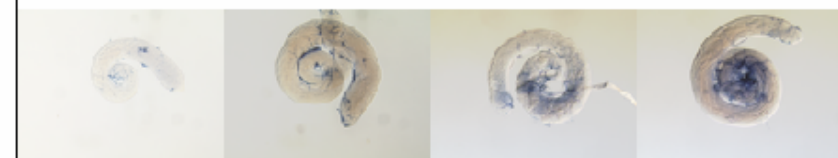
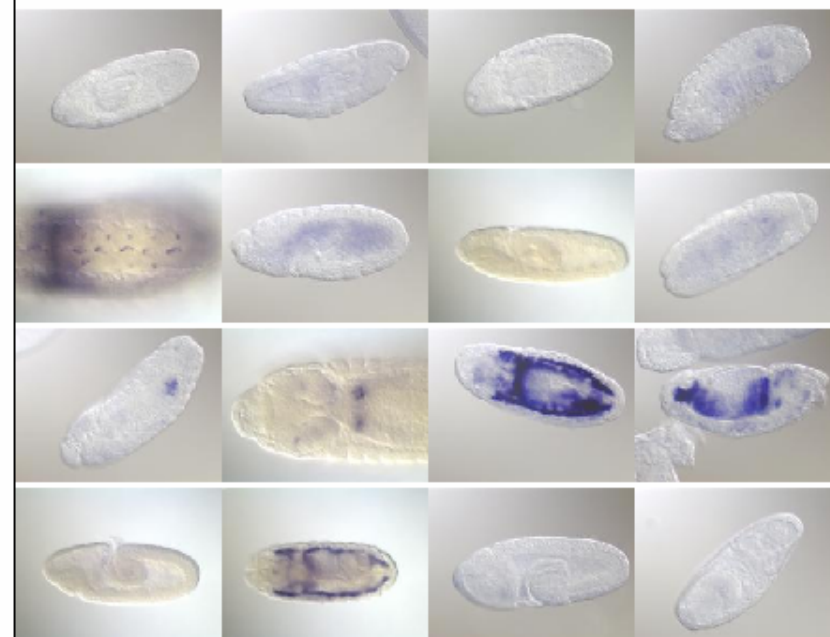


Image results from BDGP.

Found 16 results for query "http://www.fruitfly.org/CG32954".



SPARQLer Query Results - Mozilla Firefox

File Edit View History Bookmarks Tools Help

Back Forward Reload Stop Home Adblock Plus iMacros

http://localhost:8080/josekj/flyweb?query=PREFIX+dro%3A+++ Google

post to del.icio.us my del.icio.us Add to Connotea DefiningImageAccess... Dictionary.com Host antiparos.zoo.o... Start Page | D2R Ser... SPARQLer - An RDF ...

Gmail - Inbox (2) SPARQLer Query Results

## SPARQLer Query Results

g	cg
<http://id.datawebs.net/mapping_m11>	
<http://id.datawebs.net/flyweb_r2>	
<http://id.datawebs.net/flyweb_r1>	
<http://id.datawebs.net/mapping_m12>	
	<http://id.datawebs.net/flyweb_r2>
	<http://id.datawebs.net/mapping_m12>

SPARQLer Query Results - Mozilla Firefox

File Edit View History Bookmarks Tools Help

Back Forward Reload Stop Home Adblock Plus iMacros

http://localhost:8080/josekj/flyweb?query=PREFIX+dro%3A+++

Google

post to del.icio.us my del.icio.us Add to Connotea DefiningImageAccess... Dictionary.com Host antiparos.zoo.o... Start Page | D2R Ser... SPARQLer - An RDF ...

FlyWeb Applications // A SPARQLer Query Results

## SPARQLer Query Results

provenance	evidence
<http://purl.org/dc/elements/1.1/created>	"2007-12-19" ^^<http://www.w3.org/2001/XMLSchema#date>
<http://purl.org/dc/elements/1.1/hasVersion>	"1.0"
<http://purl.org/dc/elements/1.1/creator>	<http://www.datawebs.net/foaf.rdf#ibrg>
<http://www.datawebs.net/derivedFrom>	<http://id.fly-ted.org/v1.0>
<http://www.datawebs.net/derivedFrom>	<bdgp:2007-03-09>
<http://www.datawebs.net/derivedFrom>	<file:///C:/tomcat-5.0.28/Data/flybase/v3.2>



# Open questions

---

- What should be the evidence?
- Provenance in databases
- Provenance in e-Science
- Provenance in bioinformatics
- What the minimum provenance is needed for Linked Data?



# Acknowledgement

- David Shotton, Graham Klyne, and Alistair Miles



- Dr Helen White-Cooper and her research group
- JISC and BBSRC
- BDGP and FlyBase

JISC



GTGGCAATCCCTAAGATAGCCAAATATTATTATTGTTTCAGATACTCAC  
AGCAATCAATGCAGATGGAGTGGAGTCTTTGAAATCAGTGAAATT  
TCTTATCTTCACTTCACTTCACTTCACTTCACTTCACTTCACTTCACTT  
TCTTATCTTCACTTCACTTCACTTCACTTCACTTCACTTCACTTCACTT  
ATTTCTTCACTTCACTTCACTTCACTTCACTTCACTTCACTTCACTT  
AATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA  
AATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA

FlyBase





---

Thank you!

