

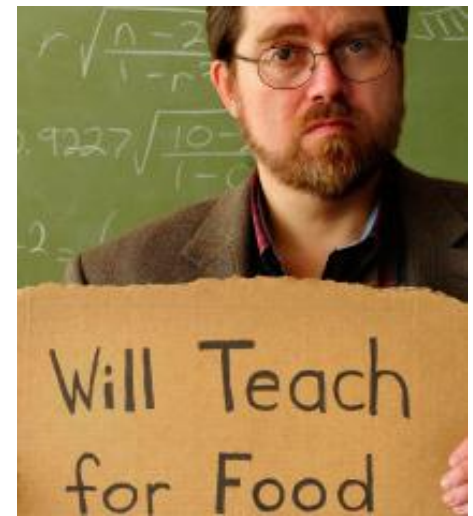
# **Developing technology that ought to be institutional infrastructure using R, FigShare and DataCite**

Björn Brembs

Universität Regensburg

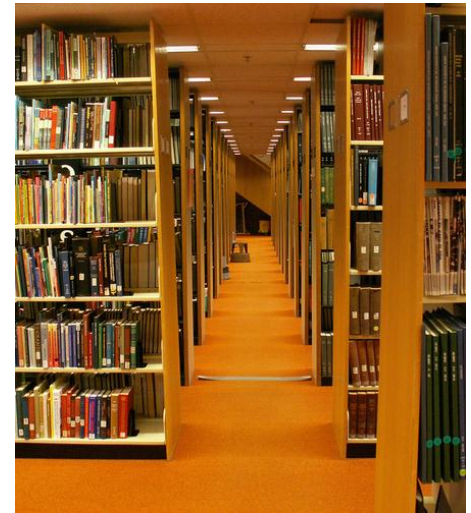
<http://brembs.net>

# SCHOLARSHIP



**Institutions produce  
publications, data and software**

# CRISIS I



## Dysfunctional scholarly literature

# Literature



...it's like the  
web in 1995!

- Limited access
- No global search
- No functional hyperlinks
- No flexible data visualization
- No submission standards
- (Almost) no statistics
- No text/data-mining
- No effective way to sort, filter and discover
- No scientific impact analysis
- No networking feature
- etc.

# CRISIS II



**Scientific data in peril**

# nature

International weekly journal of science

[Home](#)

[News & Comment](#)

[Research](#)

[Careers & Jobs](#)

[Current Issue](#)

[Archive](#)

[Audio & Video](#)

[Archive](#)

[Volume 489](#)

[Issue 7414](#)

[News](#)

[Article](#)

►► [Commendations for \*Nature\* News & Comment in the 2012 Online Media Awards](#)

NATURE | NEWS

## Databases fight funding cuts

Online tools are becoming ever more important to biology, but financial support is unstable.

**Monya Baker**

05 September 2012



Swiss Institute of  
Bioinformatics



Search UniProtKB ▼ for

Go

Clear

## ExPASy Proteomics Server

[Databases](#) [Tools](#) [Services](#) [Mirrors](#) [About](#) [Contact](#)

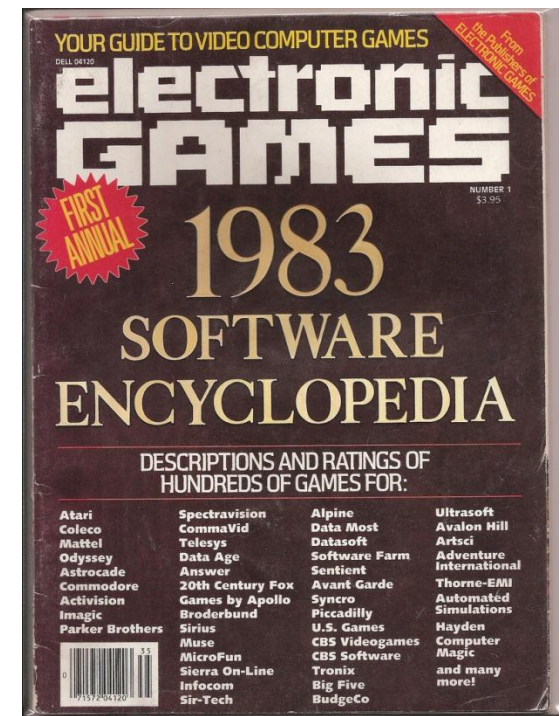
You are here: [ExPASy CH](#) > [Databases](#) > [Around UniProtKB](#)



### **SWISS-PROT should have been 10 years old in July 1996, but it may disappear on June 30, 1996**

Due to funding problems, SWISS-PROT as well as PROSITE, and the ENZYME nomenclature databases will disappear on June 30, 1996 if no solution is found before that date. The ExPASy WWW server and all services associated with it will also shut down. The distribution of the SWISS-2DPAGE database will also be discontinued. Other external databases, WWW services and software packages that depend on SWISS-PROT,

# CRISIS III



# Non-existent software archives



Science The World's Leading Journal of Original Scientific Research, Global News, and Commentary.

Science Home

Current Issue

Previous Issues

Science Express

Science Products

My Science

About the Journal

Home > Science Magazine > 2 December 2011 > Peng, 334 (6060): 1226-1227

## Article Views

- > Abstract
- > Full Text
- > Full Text (PDF)
- > Figures Only

## Article Tools

- > Leave a comment (2)
- > Save to My Folders
- > Download Citation
- > Alert Me When Article is Cited
- > Post to CiteULike

Science 2 December 2011:  
Vol. 334 no. 6060 pp. 1226-1227  
DOI: 10.1126/science.1213847

< Prev | Table of Contents | Next >



Leave a comment (2)

PERSPECTIVE

## Reproducible Research in Computational Science

Roger D. Peng

Author Affiliations

To whom correspondence should be addressed. E-mail: [rpeng@ihsph.edu](mailto:rpeng@ihsph.edu)

### ABSTRACT

Computational science has led to exciting new developments, but the nature of the work has exposed limitations in our ability to evaluate published findings. Reproducibility has the potential to serve as a minimum standard for judging scientific claims when full independent replication of a study is not possible.

# The case for open computer programs

[Darrel C. Ince](#), [Leslie Hatton](#) & [John Graham-Cumming](#)

[Affiliations](#) | [Contributions](#) | [Corresponding author](#)

*Nature* **482**, 485–488 (23 February 2012) | doi:10.1038/nature10836

Received 09 May 2011 | Accepted 05 January 2012 | Published online 22 February 2012

Scientific communication relies on evidence that cannot be entirely included in publications, but the rise of computational science has added a new layer of inaccessibility. Although it is now accepted that data should be made available on request, the current regulations regarding the availability of software are inconsistent. We argue that, with some exceptions, anything less than the release of source programs is intolerable for results that depend on computation. The vagaries of hardware, software and natural language will always ensure that exact reproducibility remains uncertain, but withholding code increases the chances that efforts to reproduce results will fail.

-  [日本語要約](#)
-  [print](#)
-  [email](#)
-  [download pdf](#)
-  [view interactive pdf](#)
-  [download citation](#)
-  [order reprints](#)
-  [rights and permissions](#)

# My Digital Utopia:

---

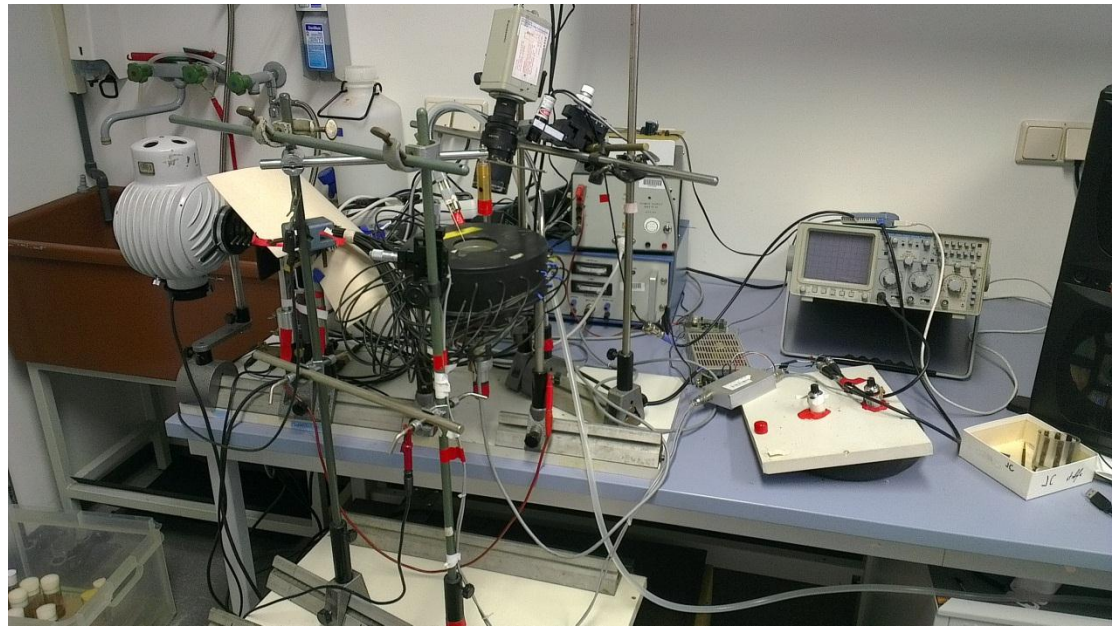
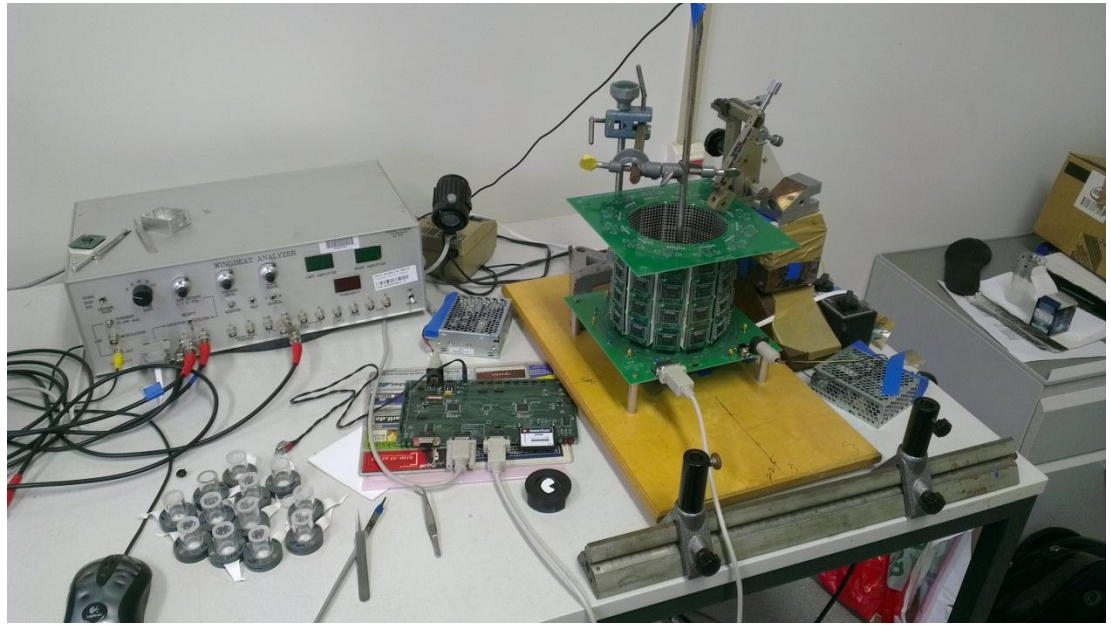
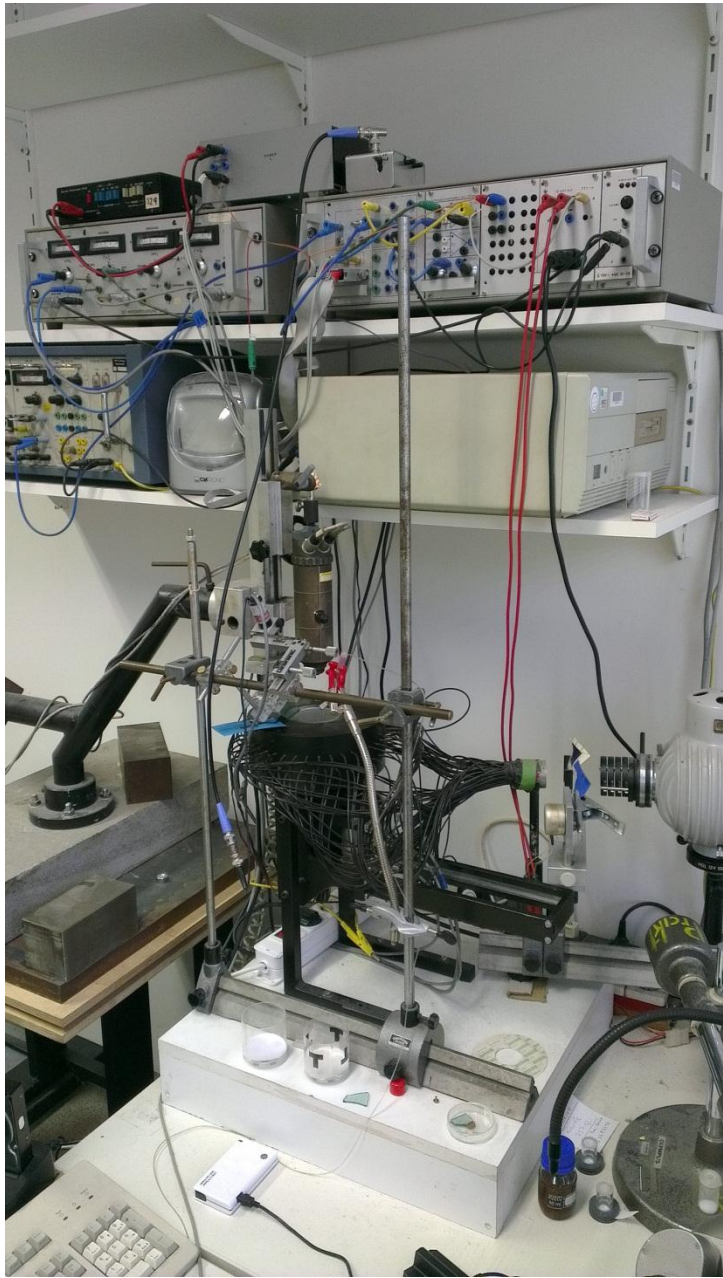
Technically feasible today (almost)

- No more corporate publishers – libraries archive everything and make it publicly accessible according to a world-wide standard
- Single semantic, decentralized database of literature, data *and* software

# Today's Digital Dystopia

- Institutional email
- Institutional webspace
- Institutional blog
- Library access card
- Open access repository
- No archiving of publications
- No archiving of software
- No archiving of data



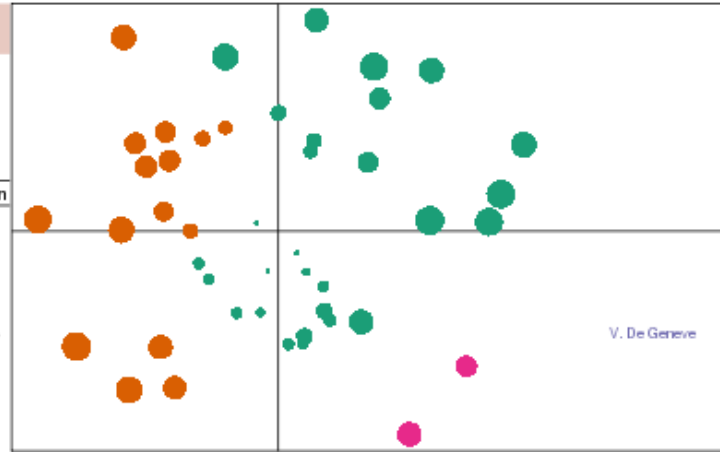
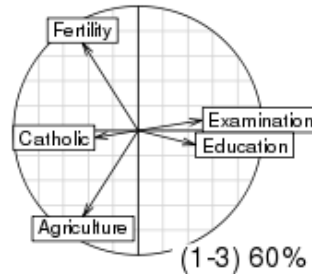


Software to control the experiment and save the data

Software to analyze and visualize the data



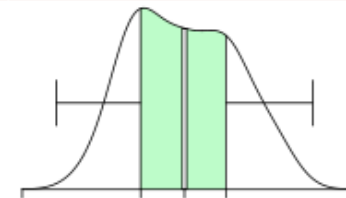
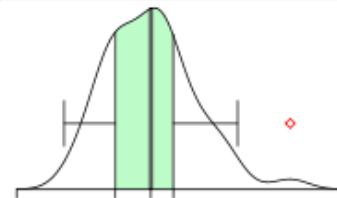
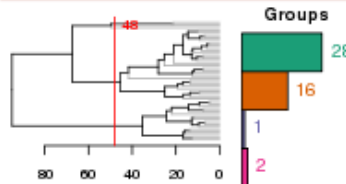
PCA 5 vars  
`princomp(x = data, cor = cor)`



Clustering 4 groups

Factor 1 [41%]

Factor 3 [19%]



V. De Geneve

About R

- [What is R?](#)
- [Contributors](#)
- [Screenshots](#)
- [What's new?](#)

Download, Packages

[CRAN](#)

R Project

- [Foundation](#)
- [Members & Donors](#)

- [Mailing Lists](#)
- [Bug Tracking](#)
- [Developer Page](#)
- [Conferences](#)
- [Search](#)

Documentation

- [Manuals](#)
- [FAQs](#)
- [The R Journal](#)
- [Wiki](#)
- [Books](#)
- [Certification](#)
- [Other](#)

Misc

- [Bioconductor](#)
- [Related Projects](#)
- [User Groups](#)
- [Links](#)

## Getting Started:

- R is a free software environment for statistical computing and graphics. It compiles and runs on a wide variety of UNIX platforms, Windows and MacOS. To [download R](#), please choose your preferred [CRAN mirror](#).
- If you have questions about R like how to download and install the software, or what the license terms are, please read our [answers to frequently asked questions](#) before you send an email.

## News :

- **R version 2.15.2** (Trick or Treat) has been released on 2012-10-26.
- [The R Journal Vol.4/1](#) is available.
- [useR! 2012](#), took place at Vanderbilt University, Nashville Tennessee, USA, June 12-15, 2012.
- [useR! 2013](#), will take place at the University of Castilla-La Mancha, Albacete, Spain, July 10-12 2013. .



# buridan.sourceforge.net

SOURCEFORGE

Find Open Source Software

[Browse](#)

[Blog](#)

[Support](#)

[Newsletters](#)

[Library](#)

[Go Parallel](#)

[Register](#)





[Log In](#)

Buridan  [brembs](#), [jwessnit](#), [lutzer](#), [thrawny](#)

[Summary](#) [Files](#) [Reviews](#) [Support](#) [Develop](#)

Home



Name ↕	Modified ↕	Size ↕	
 <a href="#">other(blueprints)</a>	2012-09-17		<a href="#">Stats</a>
 <a href="#">Analysis</a>	2012-09-17		<a href="#">Stats</a>
 <a href="#">Tracker</a>	2011-11-30		<a href="#">Stats</a>
 <a href="#">data examples</a>	2011-04-11		<a href="#">Stats</a>

Totals: 23 Items

778.9 MB

## RESEARCH ARTICLE



# Open Source Tracking and Analysis of Adult *Drosophila* Locomotion in Buridan's Paradigm with and without Visual Targets

Article

Metrics

Related Content

Comments: 1

Julien Colomb<sup>1\*</sup>, Lutz Reiter<sup>1</sup>, Jędrzej Blaszkiwicz<sup>1</sup>, Jan Wessnitzer<sup>2</sup>, Bjoern Brembs<sup>1,3</sup>

**1** FB Biologie, Chemie, Pharmazie, Institut für Biologie-Neurobiologie, Freie Universität Berlin, Berlin, Germany, **2** Institute for Perception, Action and Behaviour, School of Informatics, University of Edinburgh, Edinburgh, United Kingdom, **3** Department of Genetics, Universität Leipzig, Leipzig, Germany

## Abstract [Top](#)

### Background

**1** Insects have been among the most widely used model systems for studying the control of locomotion by nervous systems. In *Drosophila*, we implemented a simple test for locomotion: in Buridan's paradigm, flies walk back and forth between two inaccessible visual targets [1]. Until today, the lack of easily accessible tools for tracking the fly position and analyzing its trajectory has probably contributed to the slow acceptance of Buridan's paradigm.

To add a note, highlight some text. [Hide notes](#)

[Make a general comment](#)

#### Jump to

[Abstract](#)

[Introduction](#)

[Materials and Methods](#)

[Results](#)

[Discussion](#)

[Supporting Information](#)

[Acknowledgments](#)

[Author Contributions](#)

[References](#)

[View All Figures](#)

Download: [PDF](#) | [Citation](#) | [XML](#)

[Print article](#)

[EzReprint](#) New & improved!

#### Metrics [i](#)

**Total Article Views: 1,144**

Published on August 9, 2012

#### Citations

No citations found

#### Social Networks

Facebook (13)

Twitter (3)

Mendeley (8)

[More](#)

#### Related Content

##### Related Subject Categories

[Biology and life sciences](#), [Computer and information sciences](#), [Engineering and technology](#), [Physical sciences](#)

##### Related Articles on the Web

[Google Scholar](#)

[PubMed](#)

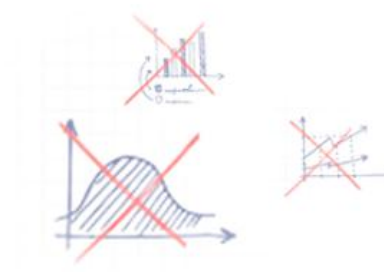
[More](#)

#### Share this Article [i](#)



[Email this article](#)

## Features



### 1GB of private space

taggable and easily filtered, your research data is better managed and easy to locate

### Unlimited public space

upload to your heart's content  
the more - the better

### Publish negative data

all published research is citable



### Upload all formats

all research outputs welcomed – images, graphs, videos, datasets



### Quick & simple upload

upload to your heart's content  
the more - the better




### Cloud based

Secure and accessible from anywhere

# SOFTWARE THAT UNDERSTANDS SCIENCE

We are an innovative technology company developing software and apps that change the way science is done.

[Visit our US site](#) [Home](#) | [Help](#) | [FAQs](#) | [Site map](#) | [Office locations](#) | [About US](#)[Doing BUSINESS](#)[Writing for MACMILLAN](#)[Working for MACMILLAN](#)[Media CENTRE](#)[How to CONTACT US](#)[Home](#) > [Media Centre](#)

## Macmillan announces launch of Digital Science



LONDON -- 7 December 2010

Macmillan Publishers Ltd. today announces the launch of Digital Science (<http://www.digital-science.com/>), a new division of the company and a new kind of scientific information enterprise. Digital Science will focus on providing world-class software tools and services to scientists, managers and funders with the ultimate aim of making research more productive through the use of technology.

The activities of Digital Science build on the reputation for editorial and technological excellence of its sister company, Nature Publishing Group (NPG), but will focus on technology-based solutions for research rather than scientific content.

The Digital Science team brings together a wide range of expertise in science and technology, and is led by Timo Hannay, former publishing director of nature.com.



# rfigshare

*An R interface to [FigShare](#)*

- Maintainer: Carl Boettiger, [cboettig](#)
- License: [CC0](#)
- Contact: Report bugs, questions, or feature requests on the [Issues Tracker](#), or get in touch with us at [info@ropensci.org](mailto:info@ropensci.org)

## Installation

---

- Note: Requires latest [httr](#) version from Github.

```
require(devtools)
install_github("httr", "hadley")
install_github("rfigshare", "ropensci")
```

## Getting Started

---

This package is still in its early development stages. Meanwhile check out [Getting Started](#) for a quick introduction to obtain your API keys and start exploring.

**However, a version is already available**





# Add the variables to your code

```
1
2 ▾ #####
3 #Inputs to enter#
4 ▾ #####
5 #variables specifying the experiment:
6
7 filenameprefix ="tubPKCixd42+-chaG80/tG80_PKCi_d42g_"
8 firstpoint =1
9 datapoints <- 51
10
11 # figshare infos
12 # create the article on the first run (if no id_test is given)
13 # after first run, return "id", you get a number to insert as a value for the id_test
14 # next runs will update the existing article (only the figure in this version)
15
16 id_test =97916
17
18
19 article_title= "d42Gal4 UPKCi experiment: with and without UASGFP and chaGal80"
20 article_description = "this is a temporary article"
21 article_type = "figure" #, "dataset" #,"media", "poster", "paper", "fileset"
22 article_tags = c("self-learning","genetics")
23 article_categories="Neuroscience"
24
25 article_visibility= "draft" #"private" "public" #
26 article_authors= c("julien colomb")
27
28 article_links="http://lab.brembs.net"
29 article_files = "T:dataforfigshare.png"
30
31 ▾ #####
32 ▾ #####
33 #start code of analysis
34 ▾ #####
```

# What to publish on figshare

```
205
206 ## Data treatment is ready for analysis and plotting:
207 source ("D:/dokumente/data/Rfiles/fs/dataanalysis3g.r") #produce a pdf with all the analysis locally
208 |
209 source ("D:/dokumente/data/Rfiles/fs/dataanalysisfigshare.r") #produce a png image with the principal results and publish it in figshare
210
```

For instance, produce an image than can be previewed (instead of a multiple page pdf)

```
1 ##### produce the output image
2 png("T:dataforfigshare.png")
3
4 layout(matrix(c(1,1,2,3), 2, 2, byrow = TRUE))
5 #
6 variables = list (as.factor (data_ordtot$genotype), data_ordtot$ind)
7 LI_mean <- tapply(data_ordtot$values, variables, mean, na.rm= TRUE)
8 LI_sem <- tapply(data_ordtot$values, variables, std.error)
9 samplesize <- as.matrix(summary (data$genotype))
10
11 bar <- barplot(LI_mean, beside=T, ylim = c(-0.7,1), main=("nocolorlearning"),
12               axisnames =F, legend.text=( paste(levels(data$genotype), "(n=",samplesize,")" ) ), args.legend = list(x = "bottomright", ncol=2) )
13 # store location of bars in bar, and
14 # plot the barplot. ylim is used to
15 # make room for the error bar later
16 arrows(bar, LI_mean + LI_sem, bar, LI_mean - LI_sem,
17        length = 0, # width of the arrowhead
18        angle = 90, # angle of the arrowhead
19        code = 3 # arrowhead in both ends
20 )
21
22 axis(1, at=seq(bar [2,1], bar [2,7],by= (length(levels(data_ordtot$genotype))+1)), tick =T,
23       labels=c("pretest","training","training","test","training","training","test"), las = 1)
24 abline(h=seq(0,1,.2),col=2, lty=3)
25 #
26 plot(data$X8_test2/10000~data$genotype, las=1, ylab="learning score", xlab="", main="score absolute during final test")
27 abline(h=0, col=2)
28 #
29 plot(data$scorerelative~data$genotype, las=3, ylab="learning score", xlab="", main="score final test - pretest")
30 abline(h=0, col=2)
31 dev.off()
```



# Add new, or update an existing article

```
55 #brembs lab account
56 options(FigshareKey = "A...g")
57 options(FigsharePrivateKey = "0...SA")
58 options(FigshareToken = "J...Q")
59 options(FigsharePrivateKey = "0...A")
60 ###end figshare info
61
62 require(rfigshare)
63 fs_auth()
64
65 ##need to create the article and get its id here: do it only once, then write the id and comment this part:
66
67 if (is.na(id_test)){
68 # article_title= "d42Ga14 UPKCi experiment: with and without UASGFP and chaGa180"
69 # article_description = "this is a temporary article"
70 # article_type = "figure" #, "dataset" #,"media", "poster", "paper", "fileset"
71 # article_tags = c("self-learning","genetics")
72 # article_categories="Neuroscience"
73 # article_files = "T:dataforfigshare.png"
74 # article_visibility= "draft" #"private" "public" #
75 # article_authors= c("julien colomb")
76 # article_links="http://lab.brembs.net"
77
78
79 id <- fs_new_article(title = article_title, description = article_description,
80                     type = article_type, tags = article_tags, categories=article_categories ,
81                     files = article_files, visibility= article_visibility, #authors = article_authors,
82                     links=article_links)
83 ##add björn as author (the "ö" leads to error on figshare at this time):
84 rfigshare::fs_add_author(article_id = id_test, author_id = 96464)
85
86 id
87 }else{
88 newfile= "T:dataforfigshare.png"
89 fs_upload(id_test, file =newfile)
90 }
91
```

# Run your script and...

My data  Browse Upload M. Brembs' lab

d42Gal4 UPKCi experiment: with and without UASGFP and chaGal80 [Edit article](#)

0 views 0 shares [click counting tool](#)

**Categories**

- Neuroscience

**Authors**

[members Brembs' lab](#)

**Tags**

- genetics
- self-learning

**Export**

- [Export to Ref. Manager](#)
- [Export to Endnote](#)
- [Export to Mendeley](#)

Share this: No sharing options for unpublished data

Cite this: d42Gal4 UPKCi experiment: with and without UASGFP and chaGal80. members Brembs' lab. [figshare](#). Retrieved 10.21. Nov 23, 2012 (GMT). <http://dx.doi.org/10.6084/m9.figshare.97917>

**Description**

this is a temporary article

**Links**

- <http://lab.brembs.net>

**Comments (0)**

Same type of experiments → same script

Default: → same categories

→ same tags

→ same authors

→ same links

→ same description

→ One complete article, in one click.

Update the figure:

Higher sample size directly published while analysed, your boss may see the results before you do! (or you may see the results of your student before they do)

Possibility to make it public and citable in one click or directly in the R code.

# Citable?

---

<http://dx.doi.org/10.6084/m9.figshare.97792>

# Citable!



Browse

Upload

Sign up

Login

PKC53e putative mutant in self learning

## Tags

• self-learning

• genetics

• figsharetests

• FBaI0197029

Cite this: PKC53e putative mutant in self learning. julien colomb, Björn Brembs.  
figshare.  
Retrieved 10:29, Dec 11, 2012 (GMT)  
<http://dx.doi.org/10.6084/m9.figshare.97792>

### Export

- [Export to Ref. Manager](#)
- [Export to Endnote](#)
- [Export to Mendeley](#)

### Description

this is a demonstration for rfigshare. This post was 99% produced in R and uploaded to figshare afterwards. This is the first step toward an automatic upload of our data on Figshare...

problems on figshare side: the file upload does not work anymore, the preview of pdf file seem not to work anymore...





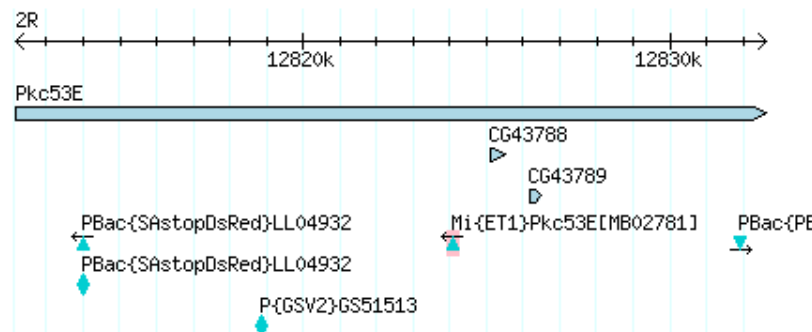
Home Tools Files Species Documents Resources News Help Archives [Jump to Gene](#) [Go](#)

Profile Manager

[Help](#)[Open All](#)[Close All](#)

## General Information

Symbol	Dmel\Pkc53E <sup>MB02781</sup>	Species	<i>D. melanogaster</i>
Name		FlyBase ID	FBal0197029
Feature type	allele	Associated gene	Dmel\Pkc53E

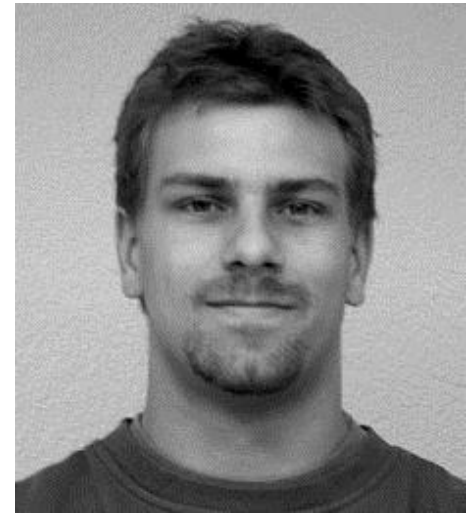
Map ( [GBrowse](#) )

## Allele class

## Mutagen

- [Recent Updates](#)
- [Nature of the Allele](#)
- [Phenotypic Data](#)
- [Interactions](#)
- [Complementation & Rescue Data](#)
- [Stocks \( 1 \)](#)
- [Notes on Origin](#)
- [External Crossreferences & Linkouts](#)
- [Synonyms & Secondary IDs \( 1 \)](#)
- [References \( 1 \)](#)

**JULIEN COLOMB**



**One person is not an  
institutional infrastructure!**