Data Without Peer: Examples of Data Peer Review in the Earth Sciences

Sarah Callaghan*
sarah.callaghan@stfc.ac.uk
@sorcha_ni

*and many others, including members of the PREPARDE, OpenAIREplus and NERC data citation and publication project teams

LCPD 2014

Why peer review data?

- Peer-review of a scientific publication is generally only applied to analysis, interpretation and conclusions, and not the underlying data.

- But if the conclusions are valid, the data must be of good quality.

- We need quality assurance of the data underlying research publications – either through peer-review or data repository checking.

- Researchers need credit for creating, managing and opening their data.

- For "Big Data" communities data checking happens as part of the sharing and archiving process, along with credit mechanisms for the data producers.

- Data journals provide academic credit for researchers in small groups, in an environment where academic status is solely based on publication record.

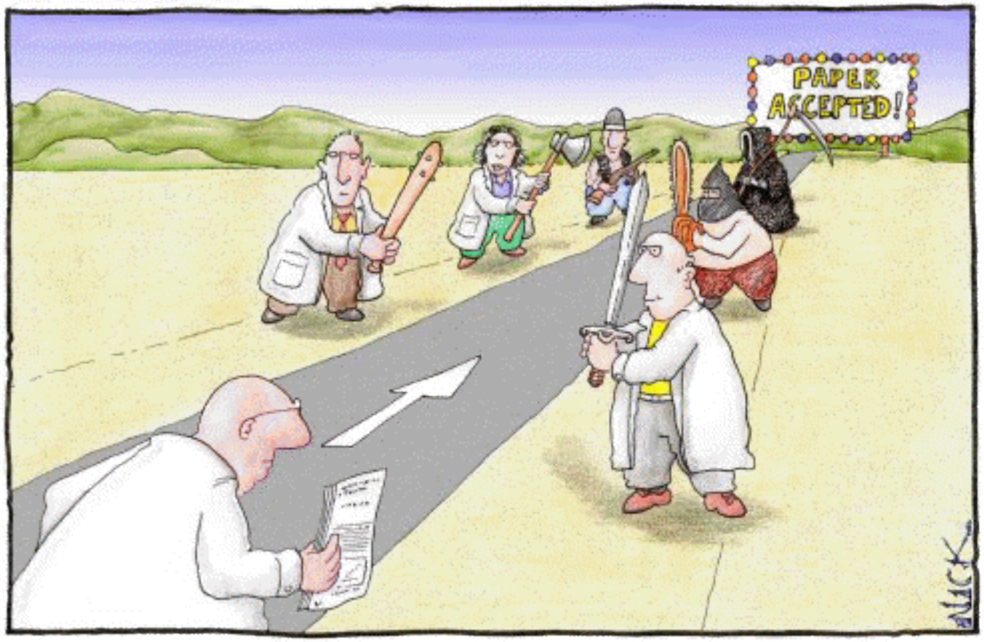
How to peer review data

The process will vary across domains.

Standardised set of questions to guide the reviewer’s thoughts.

Some portions of the review can be carried out by the journal editorial assistant, or the data repository manager hosting the data.

Many questions deal with fundamental issues regarding the accessibility of the data and understandability of the metadata

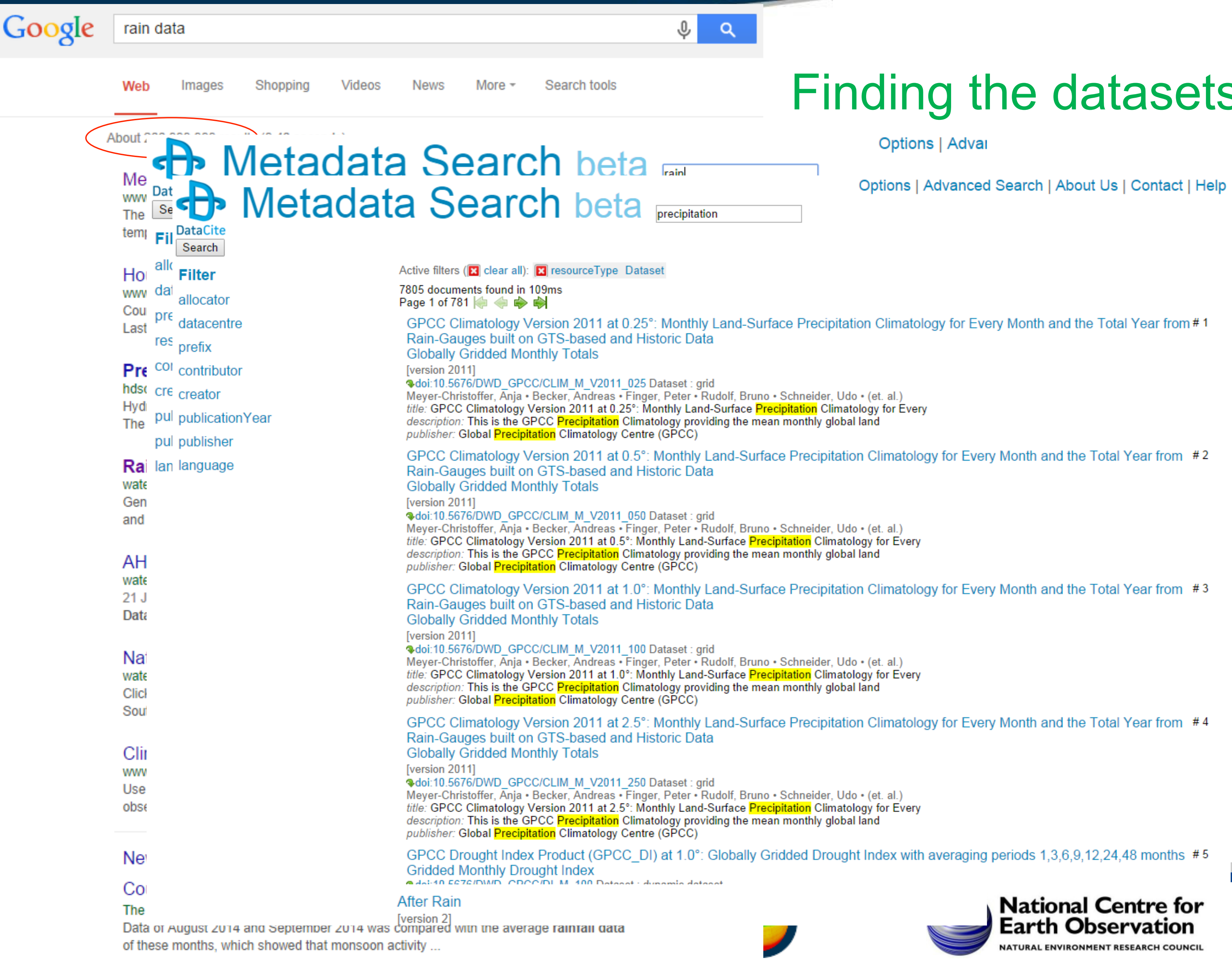


Most scientists regarded the new streamlined peer-review process as ‘quite an improvement.’

http://libguides.luc.edu/content.php?pid=5464&sid=164619

“Will I be able to use and understand this data in the future?”

Finding the datasets



Active filters (☒ clear all): ☒ resourceType Dataset

7805 documents found in 109ms
Page 1 of 781

GPCC Climatology Version 2011 at 0.25°: Monthly Land-Surface Precipitation Climatology for Every Month and the Total Year from Rain-Gauges built on GTS-based and Historic Data # 1
Globally Gridded Monthly Totals
[version 2011]
doi:10.5676/DWD_GPCC/CLIM_M_V2011_025 Dataset : grid
Meyer-Christoffer, Anja • Becker, Andreas • Finger, Peter • Rudolf, Bruno • Schneider, Udo • (et. al.)
title: GPCC Climatology Version 2011 at 0.25°: Monthly Land-Surface Precipitation Climatology for Every
description: This is the GPCC Precipitation Climatology providing the mean monthly global land
publisher: Global Precipitation Climatology Centre (GPCC)

GPCC Climatology Version 2011 at 0.5°: Monthly Land-Surface Precipitation Climatology for Every Month and the Total Year from Rain-Gauges built on GTS-based and Historic Data # 2
Globally Gridded Monthly Totals
[version 2011]
doi:10.5676/DWD_GPCC/CLIM_M_V2011_050 Dataset : grid
Meyer-Christoffer, Anja • Becker, Andreas • Finger, Peter • Rudolf, Bruno • Schneider, Udo • (et. al.)
title: GPCC Climatology Version 2011 at 0.5°: Monthly Land-Surface Precipitation Climatology for Every
description: This is the GPCC Precipitation Climatology providing the mean monthly global land
publisher: Global Precipitation Climatology Centre (GPCC)

GPCC Climatology Version 2011 at 1.0°: Monthly Land-Surface Precipitation Climatology for Every Month and the Total Year from Rain-Gauges built on GTS-based and Historic Data # 3
Globally Gridded Monthly Totals
[version 2011]
doi:10.5676/DWD_GPCC/CLIM_M_V2011_100 Dataset : grid
Meyer-Christoffer, Anja • Becker, Andreas • Finger, Peter • Rudolf, Bruno • Schneider, Udo • (et. al.)
title: GPCC Climatology Version 2011 at 1.0°: Monthly Land-Surface Precipitation Climatology for Every
description: This is the GPCC Precipitation Climatology providing the mean monthly global land
publisher: Global Precipitation Climatology Centre (GPCC)

GPCC Climatology Version 2011 at 2.5°: Monthly Land-Surface Precipitation Climatology for Every Month and the Total Year from Rain-Gauges built on GTS-based and Historic Data # 4
Globally Gridded Monthly Totals
[version 2011]
doi:10.5676/DWD_GPCC/CLIM_M_V2011_250 Dataset : grid
Meyer-Christoffer, Anja • Becker, Andreas • Finger, Peter • Rudolf, Bruno • Schneider, Udo • (et. al.)
title: GPCC Climatology Version 2011 at 2.5°: Monthly Land-Surface Precipitation Climatology for Every
description: This is the GPCC Precipitation Climatology providing the mean monthly global land
publisher: Global Precipitation Climatology Centre (GPCC)

GPCC Drought Index Product (GPCC_DI) at 1.0°: Globally Gridded Drought Index with averaging periods 1,3,6,9,12,24,48 months # 5
Gridded Monthly Drought Index
doi:10.5676/DWD_GPCC/DI_M_100 Dataset : dynamic dataset

After Rain
[version 2]

Data of August 2014 and September 2014 was compared with the average rainfall data of these months, which showed that monsoon activity ...

National Centre for Earth Observation
NATURAL ENVIRONMENT RESEARCH COUNCIL

My biases

My background is the space-time variability of rain fields

- hence choosing rain and precipitation datasets

I didn’t choose any datasets from the NERC data centres

Choices of datasets to review were made based on what I thought would be interesting examples.

- results are not statistically valid!

Data producers probably didn’t expect their data to be reviewed like this either.

g(x)≤f(x)≤h(x)
Did you read my paper on confirmation bias?
Yes, but it only proved what I already knew.

http://www.fastcompany.com/3019903/work-smart/8-subconscious-mistakes-our-brains-make-every-day-and-how-to-avoid-them

Editorial questions

- Does the dataset have a permanent identifier?
  - Yes, a DOI.
- Does it have a landing page (or README file or similar) with additional information/metadata, which allows you to determine that this is indeed the dataset you're looking for?
- Is it in an accredited/trusted repository?
- Is the dataset accessible? If not, are the terms and conditions for access clearly defined?

If the answer to any of these is “No” – dataset should be rejected without sending for review.

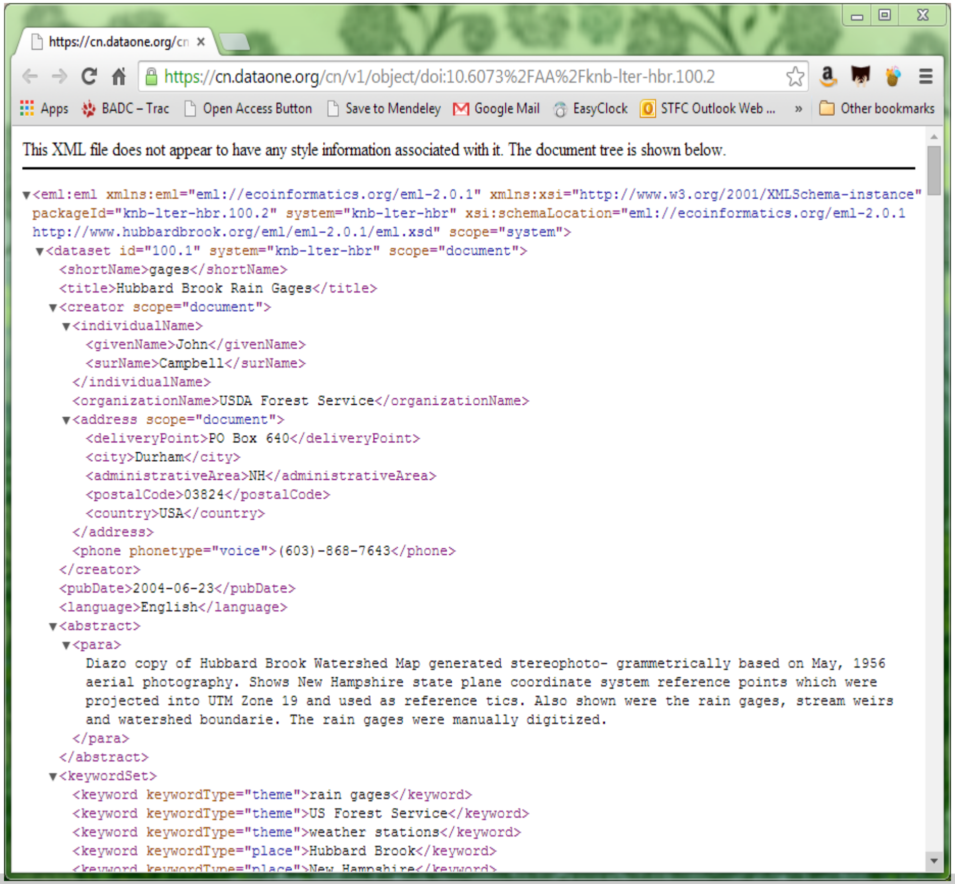
Dataset 1: Hubbard Brook Rain Gages

Citation: Campbell, John; (2004): Hubbard Brook Rain Gages; USDA Forest Service. http://dx.doi.org/10.6073/AA/KNB-LTER-HBR.100.2

Landing page?	Yes
Trusted repository?	DataONE hosting the landing page, data being held by Hubbard Brook Ecosystem Study, part of the USDA Forest Service.
Accessible dataset?	The link in the “download” section takes you to an executable file! Other links are broken. Large chunks of text dealing with Acceptable Use, Redistribution and Citation

Verdict: Revise and resubmit.

(I wouldn’t even send it to a reviewer as it is)



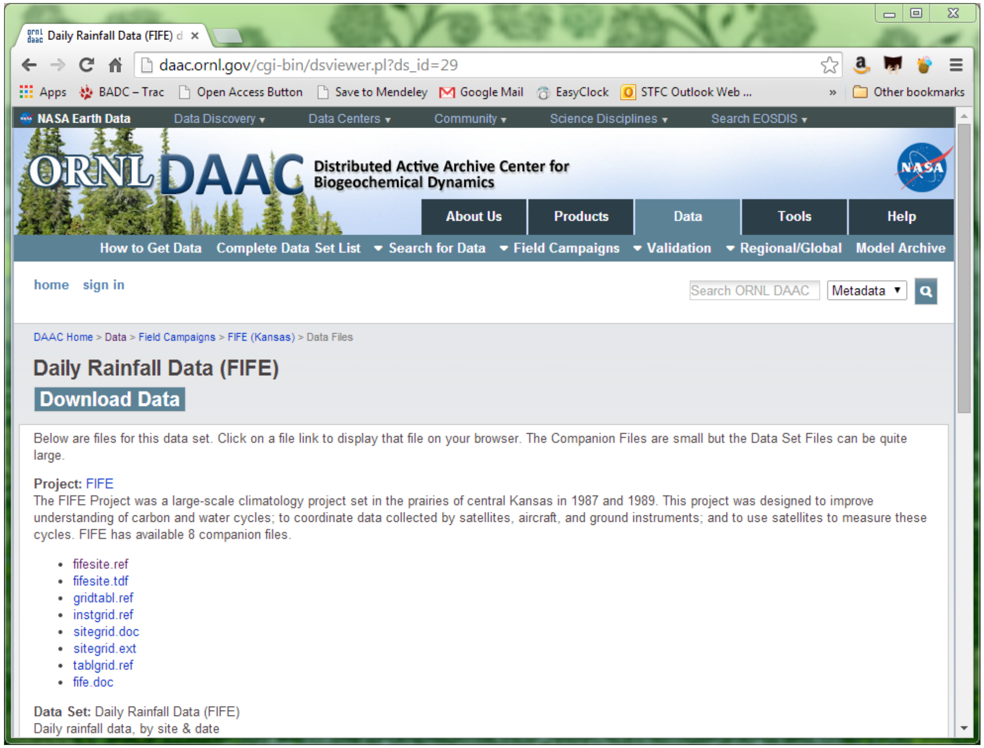
This XML file does not appear to have any style information associated with it. The document tree is shown below.

```
▼<eml:eml xmlns:eml="eml://ecoinformatics.org/eml-2.0.1" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
 packageId="knb-lter-hbr.100.2" system="knb-lter-hbr" xsi:schemaLocation="eml://ecoinformatics.org/eml-2.0.1
 http://www.hubbardbrook.org/eml/eml-2.0.1/eml.xsd" scope="system">
  ▼<dataset id="100.1" system="knb-lter-hbr" scope="document">
     <shortName>gages</shortName>
     <title>Hubbard Brook Rain Gages</title>
   ▼<creator scope="document">
     ▼<individualName>
         <givenName>John</givenName>
         <surName>Campbell</surName>
       </individualName>
       <organizationName>USDA Forest Service</organizationName>
     ▼<address scope="document">
         <deliveryPoint>PO Box 640</deliveryPoint>
         <city>Durham</city>
         <administrativeArea>NH</administrativeArea>
         <postalCode>03824</postalCode>
         <country>USA</country>
       </address>
       <phone phonetype="voice">(603)-868-7643</phone>
     </creator>
     <pubDate>2004-06-23</pubDate>
     <language>English</language>
   ▼<abstract>
     ▼<para>
         Diazo copy of Hubbard Brook Watershed Map generated stereophoto- grammetrically based on May, 1956
         aerial photography. Shows New Hampshire state plane coordinate system reference points which were
         projected into UTM Zone 19 and used as reference tics. Also shown were the rain gages, stream weirs
         and watershed boundarie. The rain gages were manually digitized.
       </para>
     </abstract>
   ▼<keywordSet>
       <keyword keywordType="theme">rain gages</keyword>
       <keyword keywordType="theme">US Forest Service</keyword>
       <keyword keywordType="theme">weather stations</keyword>
       <keyword keywordType="place">Hubbard Brook</keyword>
       <keyword keywordType="place">New Hampshire</keyword>
```

Dataset 2: Daily Rainfall Data (FIFE)

Citation: HUEMMRICH, K.F.; BRIGGS, J.M.; (1994): Daily Rainfall Data (FIFE); ORNL Distributed Active Archive Center. http://dx.doi.org/10.3334/ORNLDAAC/29

Landing page?	Yes
Trusted repository?	I know ORNL DAAC, the federation, but haven't worked with the Biogeochemical Dynamics group
Accessible dataset?	Need to sign in to download the data. Confusing list of files underneath the "Download Data" button, with the caption "Below are files for this dataset". Further down the page is "Download Data Set Files: (1.0 MBytes in 89 Files)" (with no hyperlink to click on), which seems to suggest that the files on the page aren't the data.



Daily Rainfall Data (FIFE)

Download Data

Below are files for this data set. Click on a file link to display that file on your browser. The Companion Files are small but the Data Set Files can be quite large.

Project: FIFE

The FIFE Project was a large-scale climatology project set in the prairies of central Kansas in 1987 and 1989. This project was designed to improve understanding of carbon and water cycles; to coordinate data collected by satellites, aircraft, and ground instruments; and to use satellites to measure these cycles. FIFE has available 8 companion files.

- fifesite.ref
- fifesite.tdf
- gridtabl.ref
- instgrid.ref
- sitegrid.doc
- sitegrid.ext
- tablgrid.ref
- fife.doc

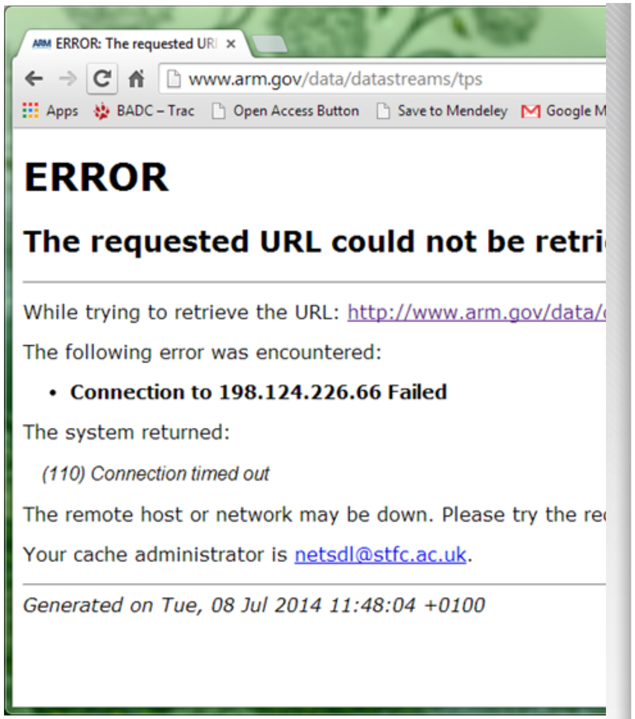
Data Set: Daily Rainfall Data (FIFE)
Daily rainfall data, by site & date

Verdict: Don't know!

Access restrictions put reviewers off.

Dataset 3: ARM: Total Precipitation Sensor

Citation: Jessica, Cherry; (2006): ARM: Total Precipitation Sensor; Not Available. http://dx.doi.org/10.5439/1025305


ARM ERROR: The requested URL
www.arm.gov/data/datastreams/tps
Apps
BADC – Trac
Open Access Button
Save to Mendeley
Google M

ERROR

The requested URL could not be retri

While trying to retrieve the URL: http://www.arm.gov/data/

The following error was encountered:

- **Connection to 198.124.226.66 Failed**

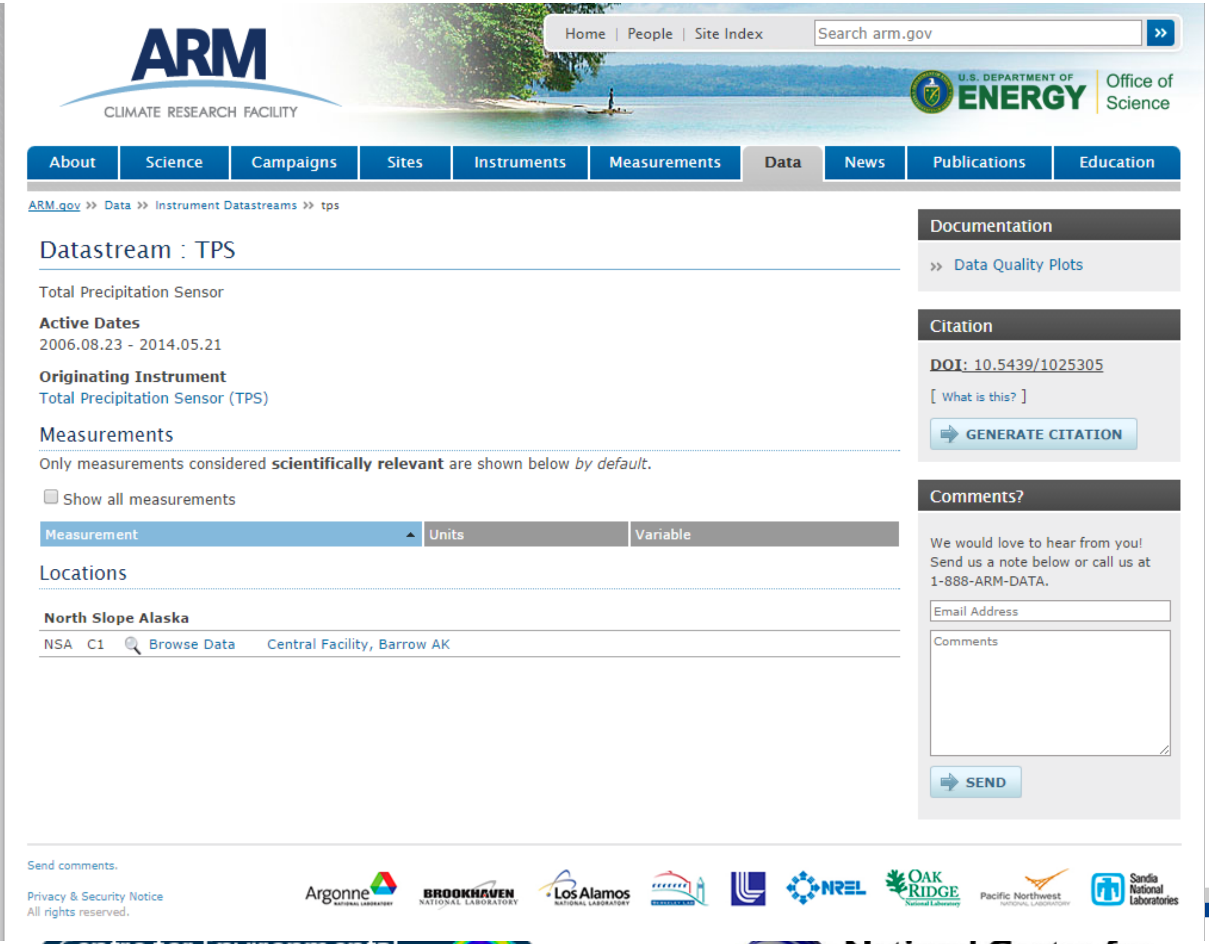
The system returned:

(110) Connection timed out

The remote host or network may be down. Please try the re

Your cache administrator is netsdl@stfc.ac.uk.

Generated on Tue, 08 Jul 2014 11:48:04 +0100


ARM
CLIMATE RESEARCH FACILITY
Home | People | Site Index
Search arm.gov
U.S. DEPARTMENT OF ENERGY
Office of Science
About
Science
Campaigns
Sites
Instruments
Measurements
Data
News
Publications
Education
ARM.gov >> Data >> Instrument Datastreams >> tps

Datastream : TPS

Total Precipitation Sensor

Active Dates
2006.08.23 - 2014.05.21

Originating Instrument
Total Precipitation Sensor (TPS)

Measurements

Only measurements considered **scientifically relevant** are shown below *by default*.

□ Show all measurements

Measurement	Units	Variable

Locations

North Slope Alaska

NSA	C1	Browse Data	Central Facility, Barrow AK

Documentation

Data Quality Plots

Citation

DOI: 10.5439/1025305

[What is this?]

GENERATE CITATION

Comments?

We would love to hear from you! Send us a note below or call us at 1-888-ARM-DATA.

Email Address

Comments

SEND

Send comments.

Privacy & Security Notice
All rights reserved.

Verdict: Reject!

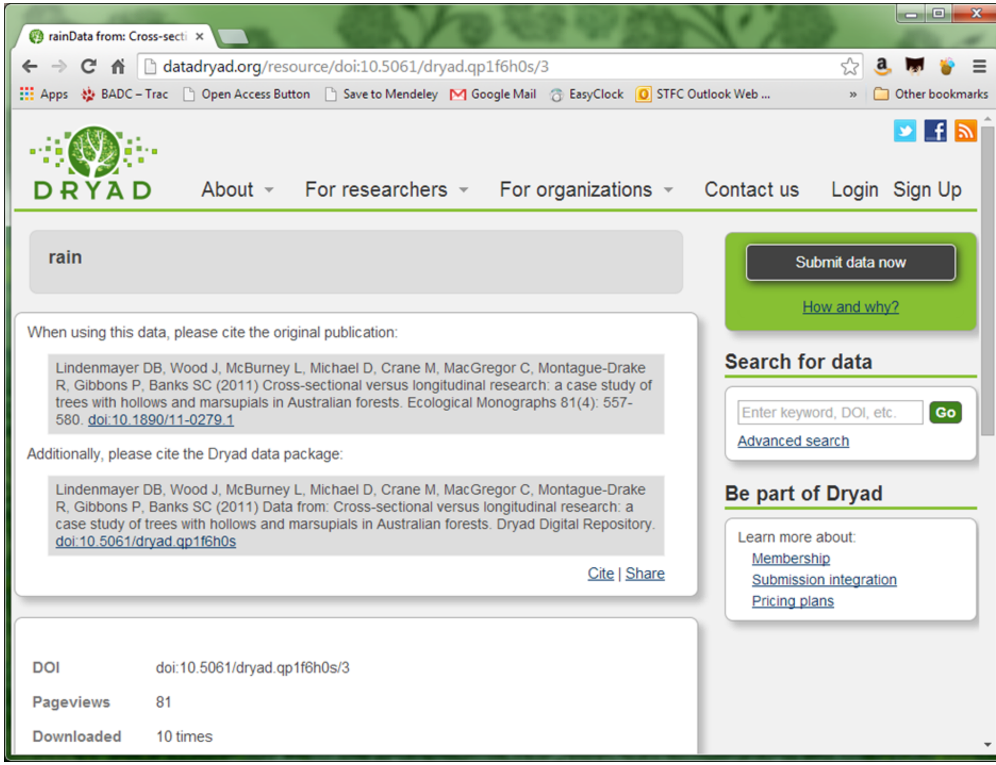
If the DOI doesn't resolve, i

for review.

Dataset 4: rain

Citation: Lindenmayer, David B.; Wood, Jeff; McBurney, Lachlan; Michael, Damian; Crane, Mason; MacGregor, Christopher; Montague-Drake, Rebecca; Gibbons, Philip; Banks, Sam C.; (2011): rain; Dryad Digital Repository. http://dx.doi.org/10.5061/DRYAD.QP1F6H0S/3

Landing page?	Yes
Trusted repository?	Yes
Accessible dataset?	Yes, both the data file rain.csv and the readme.txt file are both clearly found on the page and are easily downloadable.
Access terms and conditions appropriate?	Yes. "To the extent possible under law, the authors have waived all copyright and related or neighboring rights to this data." CC-zero and Open Data logos next to that text



rain

When using this data, please cite the original publication:

Lindenmayer DB, Wood J, McBurney L, Michael D, Crane M, MacGregor C, Montague-Drake R, Gibbons P, Banks SC (2011) Cross-sectional versus longitudinal research: a case study of trees with hollows and marsupials in Australian forests. Ecological Monographs 81(4): 557-580. doi:10.1890/11-0279.1

Additionally, please cite the Dryad data package:

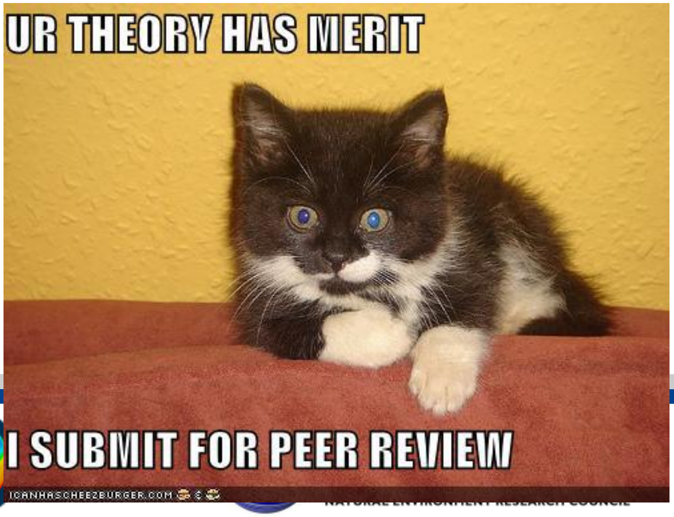
Lindenmayer DB, Wood J, McBurney L, Michael D, Crane M, MacGregor C, Montague-Drake R, Gibbons P, Banks SC (2011) Data from: Cross-sectional versus longitudinal research: a case study of trees with hollows and marsupials in Australian forests. Dryad Digital Repository. doi:10.5061/dryad.qp1f6h0s

DOI	doi:10.5061/dryad.qp1f6h0s/3
Pageviews	81
Downloaded	10 times

Dataset 4: rain

Format acceptable?	Yes - csv
Can I open the files?	Yes
Proprietary software? including version number?	Not applicable
Metadata appropriate?	The metadata is in the readme.txt file and is a simple sentence: “rain.csv contains rainfall in mm for each month at Marysville, Victoria from January 1995 to February 2009”. This is not enough metadata.
Unexplained/ non-standard acronyms in the dataset title/ metadata?	The dataset title is just “rain”, which is not very helpful at all to any potential users. On the landing page, it does show clearly that this particular dataset is, in fact, part of another larger data package

Data calibrated and calibration supplied?	Don’t know
Data flagged with explanation?	Yes – null flags, but no explanation
Metadata about how/why the data was collected?	Not in the readme file, or on the landing page itself. Maybe in the paper associated with this data package – which is paywalled
Variable names defined with units?	Not in the csv file itself, but there is a little bit of information in the readme.txt file



Dataset 4: rain

Citation: Lindenmayer, David B.; Wood, Jeff; McBurney, Lachlan; Michael, Damian; Crane, Mason; MacGregor, Christopher; Montague-Drake, Rebecca; Gibbons, Philip; Banks, Sam C.; (2011): rain; Dryad Digital Repository. http://dx.doi.org/10.5061/DRYAD.QP1F6H0S/3

Can data be reused?	Yes, but only because it’s such a simple measurement. Though providing the latitude and longitude of the site would have made it far more useful.
Data of value?	Yes – but only because it’s observational and can’t be repeated
Obvious mistakes?	No.
Data within expected ranges	Yes
Relationship between multiple data variables clear?	Not applicable

Verdict: Revise and resubmit

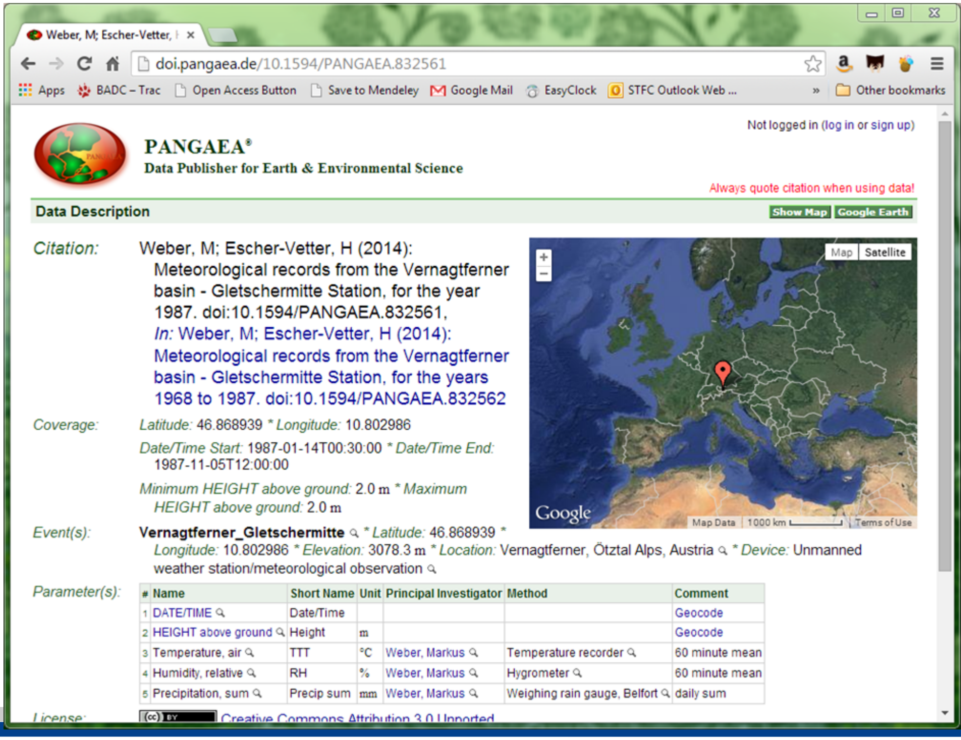
Small part of a research project not really looking at rain.
Yet data could be amalgamated with other datasets to make them more useful.
Title needs more detail, as does metadata – especially calibration, type of gauge, latitude and longitude

A Brief Pause

Dataset 5: Meteorological records from the Vernagtferner basin - Gletschermitte Station, for the year 1987

Citation: Weber, Markus; Escher-Vetter, Heidi; (2014): Meteorological records from the Vernagtferner basin - Gletschermitte Station, for the year 1987; PANGAEA - Data Publisher for Earth & Environmental Science. http://dx.doi.org/10.1594/PANGAEA.832561

Access terms and conditions appropriate?	Yes, Creative Commons Attribution 3.0 Unported
Format acceptable?	Yes. Data is provided as tab delimited text in a choice of standards.
Metadata appropriate?	Yes . Though there are gaps in the series that you' ll only see by looking at the data – it would have been good to have these gaps identified in the metadata.



PANGAEA®
Data Publisher for Earth & Environmental Science

Always quote citation when using data!

Data Description

Citation: Weber, M; Escher-Vetter, H (2014): Meteorological records from the Vernagtferner basin - Gletschermitte Station, for the year 1987. doi:10.1594/PANGAEA.832561,
In: Weber, M; Escher-Vetter, H (2014): Meteorological records from the Vernagtferner basin - Gletschermitte Station, for the years 1968 to 1987. doi:10.1594/PANGAEA.832562

Coverage: *Latitude:* 46.868939 * *Longitude:* 10.802986
Date/Time Start: 1987-01-14T00:30:00 * *Date/Time End:* 1987-11-05T12:00:00
Minimum HEIGHT above ground: 2.0 m * *Maximum HEIGHT above ground:* 2.0 m

Event(s): **Vernagtferner_Gletschermitte** * *Latitude:* 46.868939 * *Longitude:* 10.802986 * *Elevation:* 3078.3 m * *Location:* Vernagtferner, Ötztal Alps, Austria * *Device:* Unmanned weather station/meteorological observation

Parameter(s):

#	Name	Short Name	Unit	Principal Investigator	Method	Comment
1	DATE/TIME	Date/Time				Geocode
2	HEIGHT above ground	Height	m			Geocode
3	Temperature, air	TTT	°C	Weber, Markus	Temperature recorder	60 minute mean
4	Humidity, relative	RH	%	Weber, Markus	Hygrometer	60 minute mean
5	Precipitation, sum	Precip sum	mm	Weber, Markus	Weighing rain gauge, Belfort	daily sum

License: Creative Commons Attribution 3.0 Unported

Dataset 5: Meteorological records from the Vernagtferner basin - Gletschermitte Station, for the year 1987

Data calibrated and calibration supplied?	No information supplied. Gauge is given as “Weighing rain gauge, Belfort”, but it would have been helpful to give a make and model, as a Google search results in several different instruments of that type.
Data flagged with explanation?	The data isn’t flagged, which caused confusion when opening the csv file in a text editor - looked like there were no relative humidity or precipitation sum values – but they are there if the user scrolls down far enough. The html view of the first 2000 lines is helpful, as it makes it easy for the user to scroll quickly through the data.

Metadata about how/ why the data was collected?	Yes – this is a year’s worth of data from a larger dataset spanning multiple years, all at the same location: This dataset collection also provides a link to a grey literature document, also in Pangaea

Verdict: Accept

This dataset was the best documented and will be very useful!

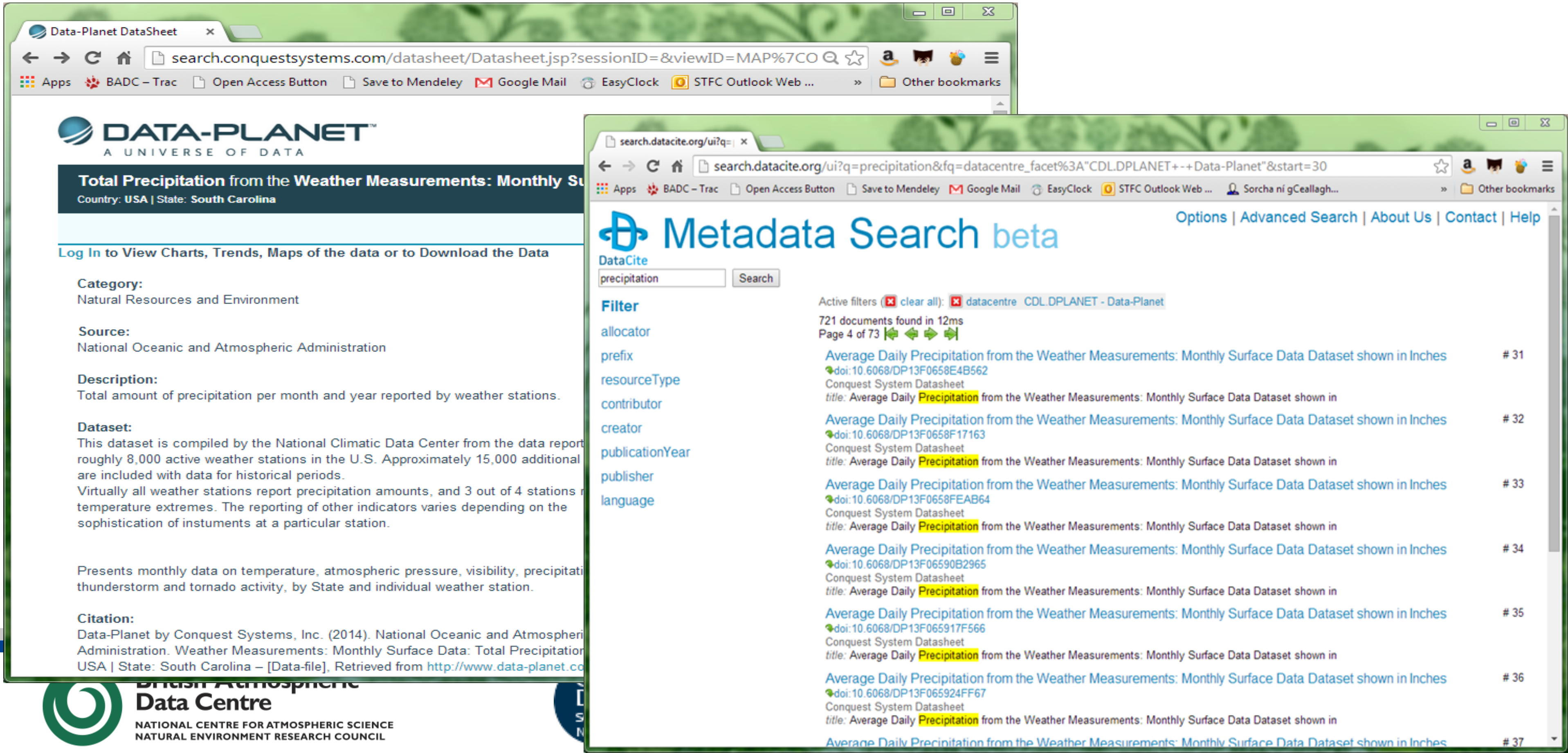
Dataset 6: National Oceanic and Atmospheric Administration. Weather Measurements: Monthly Surface Data: Total Precipitation | Country: USA | State: South Carolina – [Data-file]

Citation: Data-Planet by Conquest Systems, Inc. (2014). National Oceanic and Atmospheric Administration. Weather Measurements: Monthly Surface Data: Total Precipitation | Country: USA | State: South Carolina – [Data-file], Retrieved from http://www.data-planet.com, Viewed: July 8, 2014. Dataset-ID: 018-002-006. doi:10.6068/ DP143A169EBCB2

Or (DataCite citation)

Conquest System Datasheet; (2013): Average Daily Precipitation from the Weather Measurements: Monthly Surface Data Dataset shown in Inches; Conquest Systems, Inc.. http://dx.doi.org/10.6068/DP143A169EBCB2

Dataset 6: National Oceanic and Atmospheric Administration. Weather Measurements: Monthly Surface Data: Total Precipitation | Country: USA | State: South Carolina – [Data-file]



Data-Planet DataSheet
search.conquestsystems.com/datasheet/Datasheet.jsp?sessionID=&viewID=MAP%7CO
Apps BADC – Trac Open Access Button Save to Mendeley Google Mail EasyClock STFC Outlook Web ... Other bookmarks
DATA-PLANET
A UNIVERSE OF DATA

Total Precipitation from the **Weather Measurements: Monthly Su**

Country: **USA** | State: **South Carolina**

Log In to View Charts, Trends, Maps of the data or to Download the Data

Category:
Natural Resources and Environment

Source:
National Oceanic and Atmospheric Administration

Description:
Total amount of precipitation per month and year reported by weather stations.

Dataset:
This dataset is compiled by the National Climatic Data Center from the data report roughly 8,000 active weather stations in the U.S. Approximately 15,000 additional are included with data for historical periods.
Virtually all weather stations report precipitation amounts, and 3 out of 4 stations r temperature extremes. The reporting of other indicators varies depending on the sophistication of instuments at a particular station.

Presents monthly data on temperature, atmospheric pressure, visibility, precipitati thunderstorm and tornado activity, by State and individual weather station.

Citation:
Data-Planet by Conquest Systems, Inc. (2014). National Oceanic and Atmospheri Administration. Weather Measurements: Monthly Surface Data: Total Precipitation USA | State: South Carolina – [Data-file], Retrieved from http://www.data-planet.co

search.datacite.org/ui?q=
search.datacite.org/ui?q=precipitation&fq=datacentre_facet%3A"CDL.DPLANET+-+Data-Planet"&start=30
Apps BADC – Trac Open Access Button Save to Mendeley Google Mail EasyClock STFC Outlook Web ... Sorcha ní gCeallagh... Other bookmarks
Options | Advanced Search | About Us | Contact | Help
DataCite
Metadata Search beta
precipitation
Search
Filter
allocator
prefix
resourceType
contributor
creator
publicationYear
publisher
language

Active filters (☒ clear all): ☒ datacentre CDL.DPLANET - Data-Planet

721 documents found in 12ms
Page 4 of 73

Average Daily Precipitation from the Weather Measurements: Monthly Surface Data Dataset shown in Inches # 31
doi:10.6068/DP13F0658E4B562
Conquest System Datasheet
title: Average Daily Precipitation from the Weather Measurements: Monthly Surface Data Dataset shown in

Average Daily Precipitation from the Weather Measurements: Monthly Surface Data Dataset shown in Inches # 32
doi:10.6068/DP13F0658F17163
Conquest System Datasheet
title: Average Daily Precipitation from the Weather Measurements: Monthly Surface Data Dataset shown in

Average Daily Precipitation from the Weather Measurements: Monthly Surface Data Dataset shown in Inches # 33
doi:10.6068/DP13F0658FEAB64
Conquest System Datasheet
title: Average Daily Precipitation from the Weather Measurements: Monthly Surface Data Dataset shown in

Average Daily Precipitation from the Weather Measurements: Monthly Surface Data Dataset shown in Inches # 34
doi:10.6068/DP13F06590B2965
Conquest System Datasheet
title: Average Daily Precipitation from the Weather Measurements: Monthly Surface Data Dataset shown in

Average Daily Precipitation from the Weather Measurements: Monthly Surface Data Dataset shown in Inches # 35
doi:10.6068/DP13F065917F566
Conquest System Datasheet
title: Average Daily Precipitation from the Weather Measurements: Monthly Surface Data Dataset shown in

Average Daily Precipitation from the Weather Measurements: Monthly Surface Data Dataset shown in Inches # 36
doi:10.6068/DP13F065924FF67
Conquest System Datasheet
title: Average Daily Precipitation from the Weather Measurements: Monthly Surface Data Dataset shown in

Average Daily Precipitation from the Weather Measurements: Monthly Surface Data Dataset shown in Inches # 37

Dataset 6: National Oceanic and Atmospheric Administration. Weather Measurements: Monthly Surface Data: Total Precipitation | Country: USA | State: South Carolina – [Data-file]

Trusted repository?	Unknown
Accessible dataset? Access terms and conditions appropriate?	No and no. The text at the top of the page says “Log In to View Charts, Trends, Maps of the data or to Download the Data”. Clicking on the login link takes you to a login page where you can login if you have an existing account. No information on that page about how to register a new account, or even a link to a help page. Top level page of the site gives you a link for FAQs, where you learn that it’s a subscriber only platform, where the cost “varies according to type of institution and size of user population”. http://homepage.data-planet.com/faq

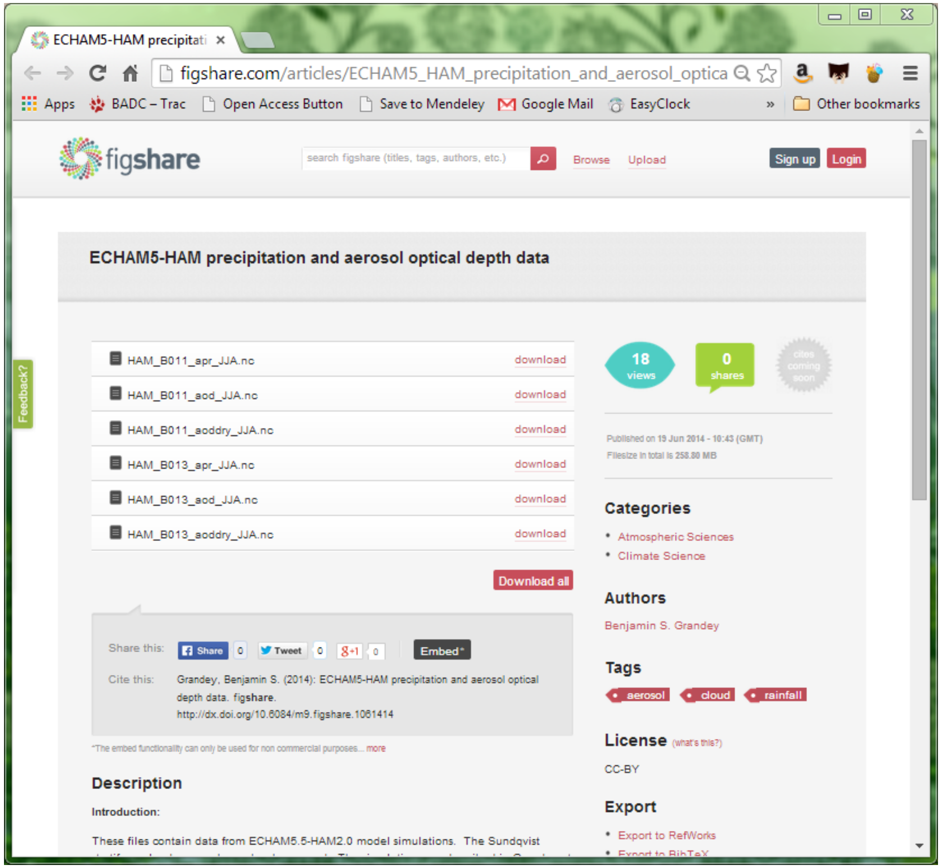
Verdict: Reject

Don’t even send out for review

Dataset 7: ECHAM5-HAM precipitation and aerosol optical depth data

Citation: Benjamin S. Grandey; (2014): ECHAM5-HAM precipitation and aerosol optical depth data; Figshare. http://dx.doi.org/10.6084/M9.FIGSHARE.1061414

Format acceptable?	*.nc files – assumed (rightly) to be netcdf, but not explicitly stated. Big files, so a warning would be useful before download
Proprietary software? including version number?	No information provided on dataset landing page
Metadata appropriate?	Metadata on landing page only gives a short outline of what the data is, and the naming conventions of the files. Metadata in the headers of the files, standard procedure for netcdf, which gives variable names, units etc.



ECHAM5-HAM precipitation and aerosol optical depth data

HAM_B011_apr_JJA.nc	download
HAM_B011_aod_JJA.nc	download
HAM_B011_aoddry_JJA.nc	download
HAM_B013_apr_JJA.nc	download
HAM_B013_aod_JJA.nc	download
HAM_B013_aoddry_JJA.nc	download

Published on 19 Jun 2014 - 10:43 (GMT)
Filesize in total is 258.80 MB

Cite this: Grandey, Benjamin S. (2014): ECHAM5-HAM precipitation and aerosol optical depth data. figshare. http://dx.doi.org/10.6084/m9.figshare.1061414

*The embed functionality can only be used for non commercial purposes... more

Description

Introduction:

These files contain data from ECHAM5.5-HAM2.0 model simulations. The Sundqvist

Categories

- Atmospheric Sciences
- Climate Science

Authors

Benjamin S. Grandey

Tags

aerosol cloud rainfall

License (what's this?)

CC-BY

Export

- Export to RefWorks
- Export to BibTeX

Dataset 7: ECHAM5-HAM precipitation and aerosol optical depth data

Unexplained/non-standard acronyms in the dataset title/ metadata?	ECHAM5.5-HAM2.0 is the model name. Citations on the dataset page to the model used and the Sundqvist stratiform cloud cover scheme would have been helpful.
Metadata about how/why the data was collected?	Only in the related paper - which is open access.
Can data be reused?	Yes – only because of the in-file metadata in the netcdf files.
Data of value?	Model data can be rerun to reproduce it. Making it available allows users to check and verify the linked papers conclusions more easily.
Obvious mistakes? Expected ranges?	Hard to tell due to no easy to use viewer.

Verdict: Accept

Would have been easier to review if I was more of a climate modeller.

Metadata on the landing page wasn't really enough to allow reuse.

In-file metadata is good, but requires the user to know what the file is and how to open it.

Conclusions (on a dataset level)

Dataset	Conclusion
Hubbard Brook rain gauges	Landing pages need to be human as well as machine readable.
Daily Rainfall Data (FIFE)	Access controls (especially registering to view datasets) put reviewers off.
ARM: Total Precipitation Sensor	If the DOI doesn't resolve, don't bother sending it to the reviewer.
rain	Relying on published papers to provide context and metadata for data doesn't work if they're behind a paywall.
Meteorological records from the Vernagtferner basin - Gletschermitte Station, for the year 1987	Good metadata makes reviewing so much easier. Linking the datasets to their parent collection and providing access to grey literature (project documents) also supports the reuse of the data.

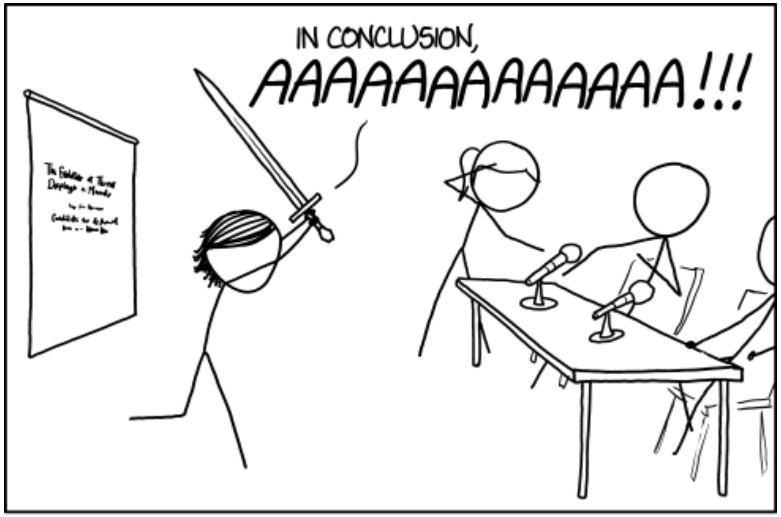
Conclusions (on a dataset level)

Dataset	Conclusion
National Oceanic and Atmospheric Administration. Weather Measurements: Monthly Surface Data: Total Precipitation | Country: USA | State: South Carolina – [Data-file]	Be consistent with citations and dataset metadata.
ECHAM5-HAM precipitation and aerosol optical depth data	In-file metadata is very helpful, but the dataset needs metadata about the file formats available before the user even gets to the data files.

Conclusions (overall)

Problems aren’t necessarily with the datasets themselves, but the way the repository makes the data available (or not)

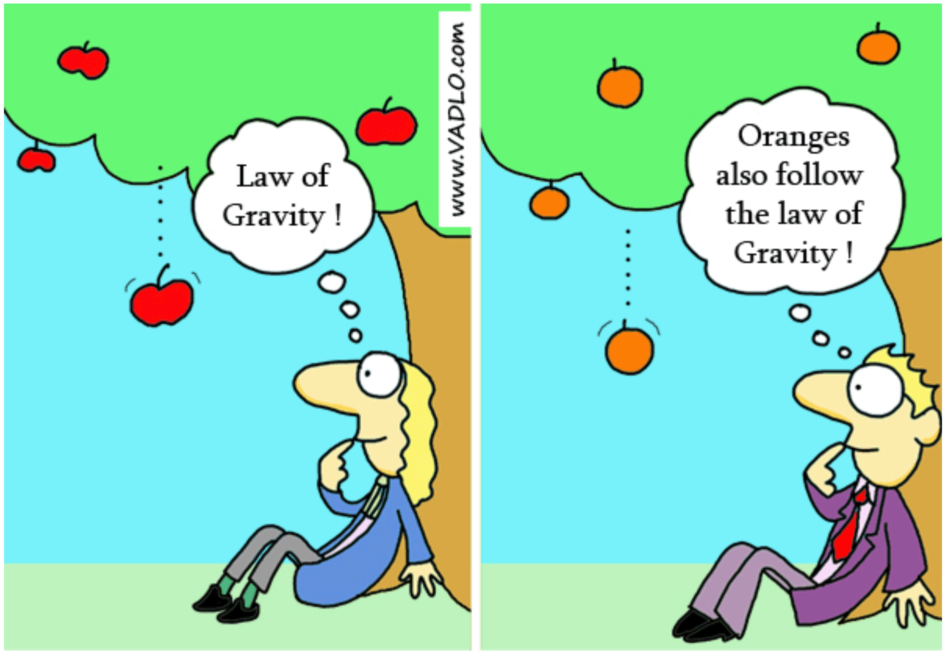
- Accessibility is a major issue – if a dataset isn’t open to the reviewer, then it’s not possible for it to be reviewed.
  - Even minor blocks could put reviewers off.
  - If important metadata for the dataset is locked in a paper behind a paywall, then that reduces the usability of the dataset.
- Human-readable metadata is critical.
  - Peer-review won’t be done by machines any time soon, so the dataset’s metadata has to be open and easily readable by human reviewers..
- Linking from the dataset landing pages to other sources of metadata is helpful, but these links need to be maintained.



THE BEST THESIS DEFENSE IS A GOOD THESIS OFFENSE.

http://xkcd.com/1403/

How to quantify impact?



High Impact Paper Low Impact Paper

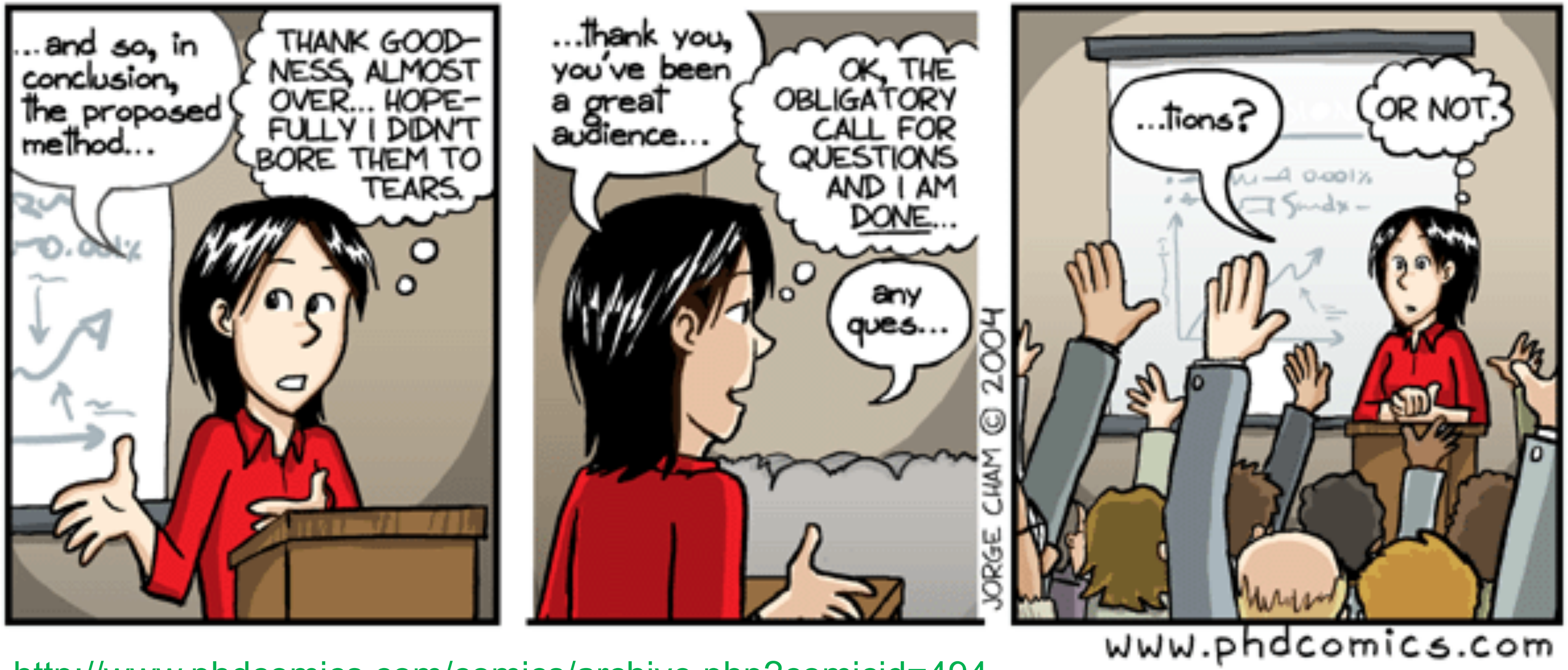
The impact of a dataset can only be determined by time!

•Would an 18th century ship's captain have realised how important their logs of meteorological measurements would be to climate scientists in the 21st century?

But we can know that if a dataset isn't useable now, it's going to be no use in the future.

Impact needs usability!

Thanks for your attention


...and so, in conclusion, the proposed method...
THANK GOODNESS, ALMOST OVER... HOPEFULLY I DIDN'T BORE THEM TO TEARS.
...thank you, you've been a great audience...
OK, THE OBLIGATORY CALL FOR QUESTIONS AND I AM DONE...
any ques...
JORGE CHAM © 2004
...tions?
OR NOT.
www.phdcomics.com

http://www.phdcomics.com/comics/archive.php?comicid=494

sarah.callaghan@stfc.ac.uk
@sorcha_ni
http://citingbytes.blogspot.co.uk/

Work funded by the European Commission as part of the project OpenAIREplus (FP7-INFRA-2011-2, Grant Agreement no. 283595)

OH OH OH
I'Z NO DEH ANSWER
ICANHASCHEEZBURGER.COM

Idea Kitteh
'ding'
has brite idea
ICANHASCHEEZBURGER.COM