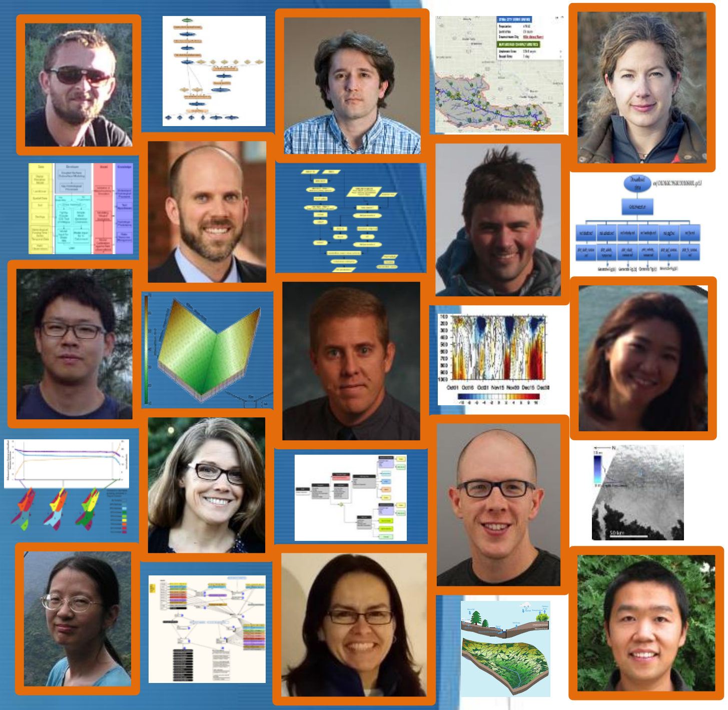
The Geoscience Paper of the Future:

Practical Guidelines for Adopting Digital Scholarship, Reproducible Research, and Open Science



Yolanda Gil

Information Sciences Institute and
Department of Computer Science
University of Southern California

gil@isi.edu

GEOSCIENCE PAPERS OF THE FUTURE
Onto Soft

Onto Soft
ICER-1440323
ICER-1343800

InGeO
Integrated Geoscience Observatory
ICER-1541057

NSF

EarthCube

http://www.ontosoft.org/gpf

Please credit as: Gil, Yolanda (Ed.) The Scientific Paper of the Future: OntoSoft Training. zenodo. http://dx.doi.org/10.5281/zenodo.159206 Retrieved 04:21, Oct 3, 2016 (GMT). October 2016.

Acknowledgments

EarthCube!

OntoSoft

ICER-1440323
ICER-1343800

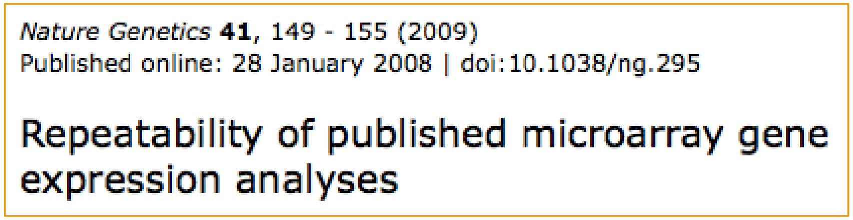
InGeO
Integrated
Geoscience
Observatory

ICER-1541057

- The Scientific Paper of the Future training materials were developed and edited by Yolanda Gil (USC), based on the OntoSoft Geoscience Paper of the Future (GPF) training materials with contributions from the OntoSoft team including Chris Duffy (PSU), Daniel Garijo (USC), Chris Mattmann (JPL), Scott Peckham (CU), Ji-Hyun Oh (USC), Varun Ratnakar (USC), Erin Robinson (ESIP)
- The OntoSoft training materials were significantly improved through input from GPF pioneers Cedric David (JPL), Ibrahim Demir (UI), Bakinam Essawy (UV), Robinson W. Fulweiler (BU), Jon Goodall (UV), Leif Karlstrom (UO), Kyo Lee (JPL), Heath Mills (UH), Suzanne Pierce (UT), Allen Pope (CU), Mimi Tzeng (DISL), Karan Venayagamoorthy (CSU), Sandra Villamizar (UC), and Xuan Yu (UD)
- Thank you to Ruth Duerr (NSIDC), James Howison (UT), Matt Jones (UCSB), Lisa Kempler (Matworks), Kerstin Lehnert (LDEO), Matt Meyernick (NCAR), Gail Clement (CalTech), and Greg Wilson (Software Carpentry) for feedback on best practices
- Thank you also to the many people that have taken the training and asked hard questions
- We are grateful for the support of the National Science Foundation and the EarthCube program

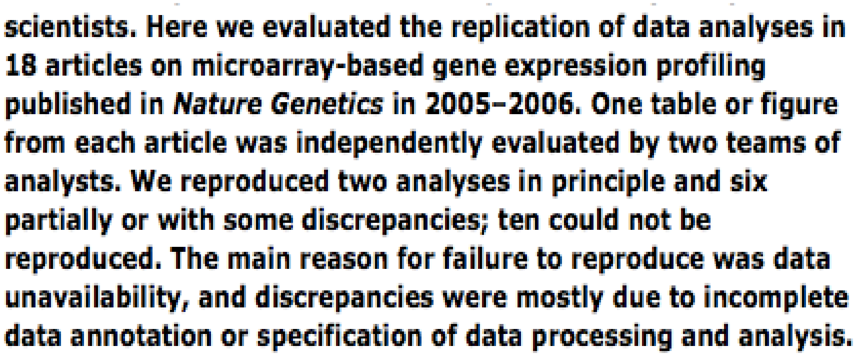
Problems with Current Practice

- Data is often not made available in publications
  - Lack of reproducibility



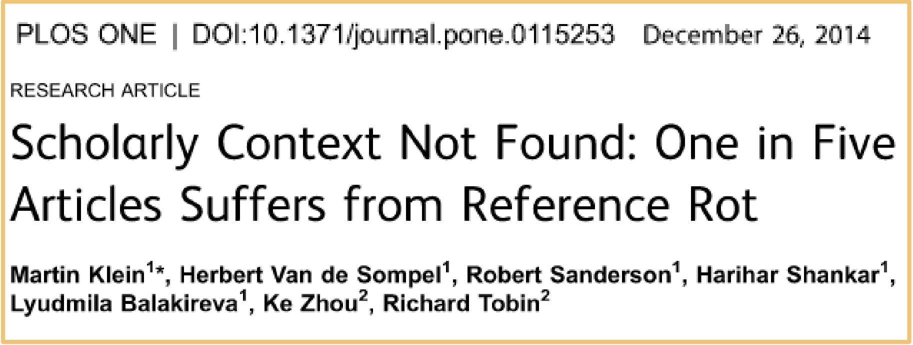
Nature Genetics **41**, 149 - 155 (2009)
Published online: 28 January 2008 | doi:10.1038/ng.295

Repeatability of published microarray gene expression analyses



scientists. Here we evaluated the replication of data analyses in 18 articles on microarray-based gene expression profiling published in *Nature Genetics* in 2005–2006. One table or figure from each article was independently evaluated by two teams of analysts. We reproduced two analyses in principle and six partially or with some discrepancies; ten could not be reproduced. The main reason for failure to reproduce was data unavailability, and discrepancies were mostly due to incomplete data annotation or specification of data processing and analysis.

- Data made available through URLs that are not persistent
  - URL does not resolve (i.e., ''rotten'')

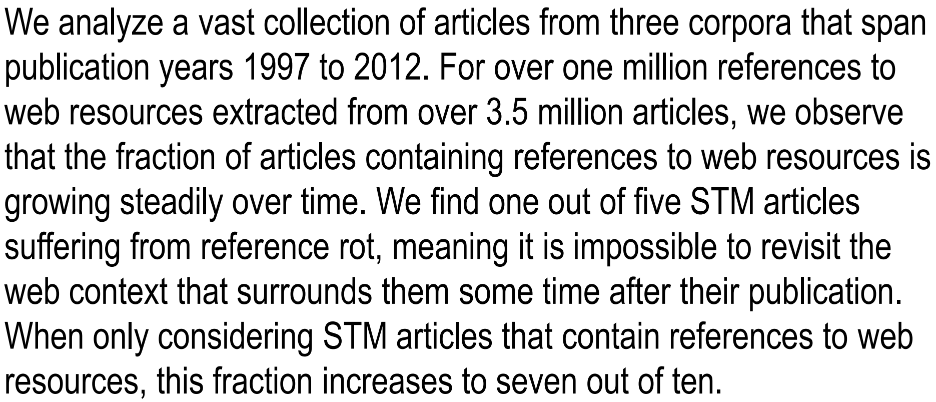


PLOS ONE | DOI:10.1371/journal.pone.0115253 December 26, 2014

RESEARCH ARTICLE

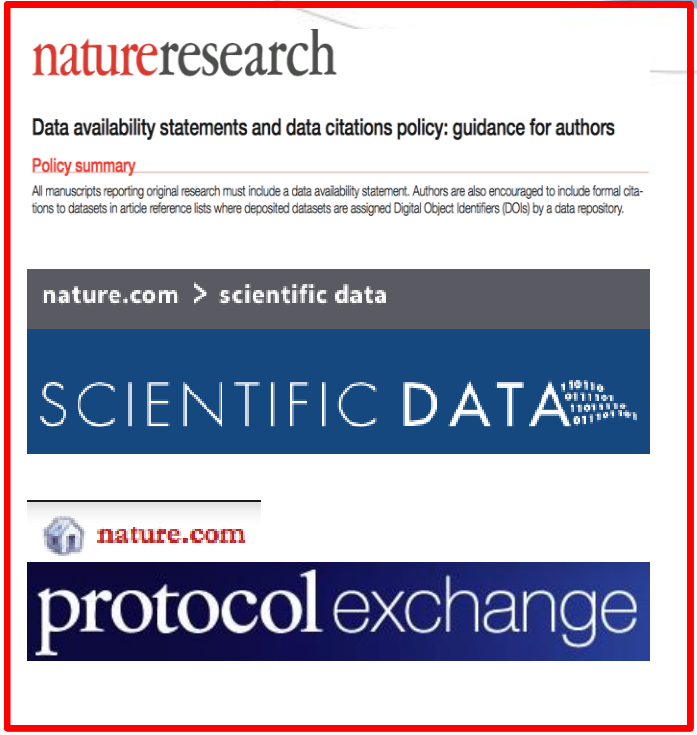
Scholarly Context Not Found: One in Five Articles Suffers from Reference Rot

Martin Klein[1]*, Herbert Van de Sompel[1], Robert Sanderson[1], Harihar Shankar[1], Lyudmila Balakireva[1], Ke Zhou[2], Richard Tobin[2]



We analyze a vast collection of articles from three corpora that span publication years 1997 to 2012. For over one million references to web resources extracted from over 3.5 million articles, we observe that the fraction of articles containing references to web resources is growing steadily over time. We find one out of five STM articles suffering from reference rot, meaning it is impossible to revisit the web context that surrounds them some time after their publication. When only considering STM articles that contain references to web resources, this fraction increases to seven out of ten.

Publishers Are Changing: Guidelines for Authors



natureresearch

Data availability statements and data citations policy: guidance for authors

Policy summary

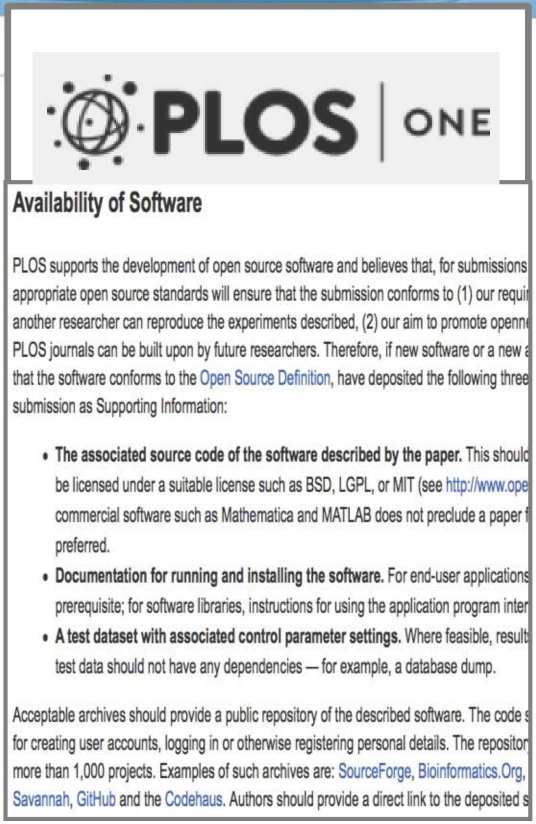
All manuscripts reporting original research must include a data availability statement. Authors are also encouraged to include formal citations to datasets in article reference lists where deposited datasets are assigned Digital Object Identifiers (DOIs) by a data repository.

nature.com > scientific data

SCIENTIFIC DATA

nature.com

protocol exchange



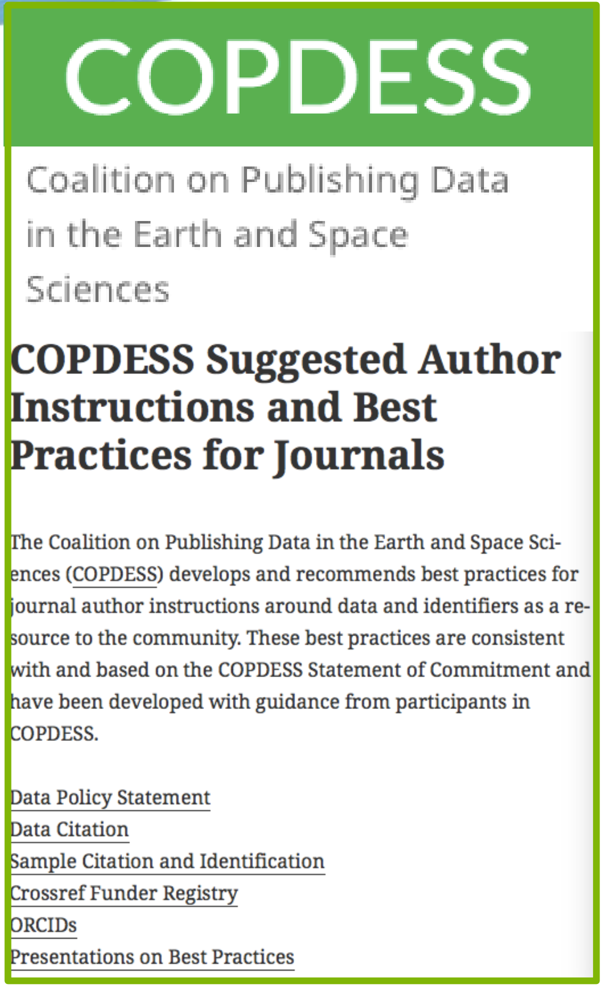
PLOS | ONE

Availability of Software

PLOS supports the development of open source software and believes that, for submissions appropriate open source standards will ensure that the submission conforms to (1) our requi another researcher can reproduce the experiments described, (2) our aim to promote openn PLOS journals can be built upon by future researchers. Therefore, if new software or a new a that the software conforms to the Open Source Definition, have deposited the following three submission as Supporting Information:

- **The associated source code of the software described by the paper.** This should be licensed under a suitable license such as BSD, LGPL, or MIT (see http://www.ope commercial software such as Mathematica and MATLAB does not preclude a paper f preferred.
- **Documentation for running and installing the software.** For end-user applications prerequisite; for software libraries, instructions for using the application program inter
- **A test dataset with associated control parameter settings.** Where feasible, result test data should not have any dependencies — for example, a database dump.

Acceptable archives should provide a public repository of the described software. The code s for creating user accounts, logging in or otherwise registering personal details. The repositor more than 1,000 projects. Examples of such archives are: SourceForge, Bioinformatics.Org, Savannah, GitHub and the Codehaus. Authors should provide a direct link to the deposited s



COPDESS

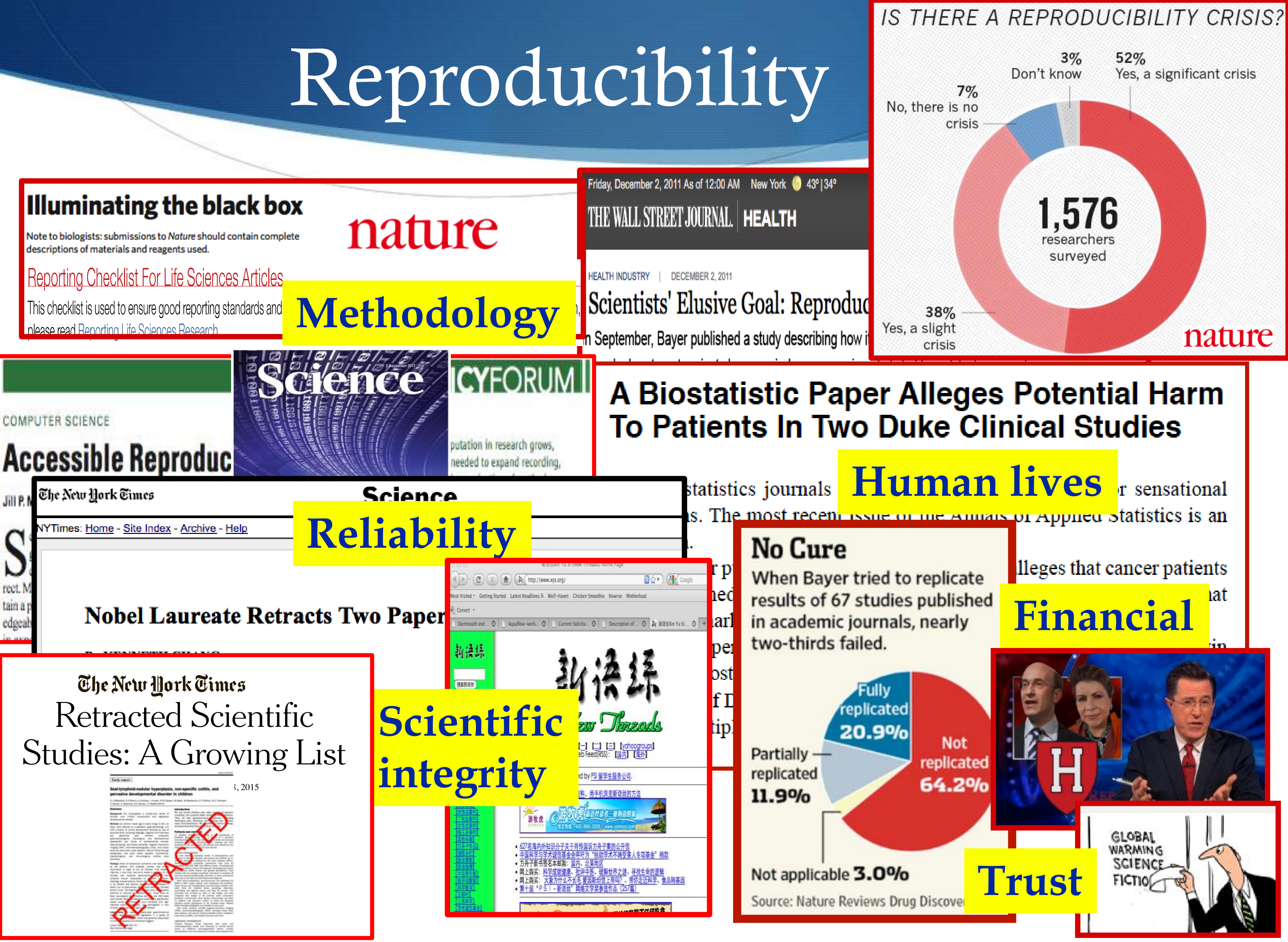
Coalition on Publishing Data in the Earth and Space Sciences

COPDESS Suggested Author Instructions and Best Practices for Journals

The Coalition on Publishing Data in the Earth and Space Sciences (COPDESS) develops and recommends best practices for journal author instructions around data and identifiers as a resource to the community. These best practices are consistent with and based on the COPDESS Statement of Commitment and have been developed with guidance from participants in COPDESS.

Data Policy Statement
Data Citation
Sample Citation and Identification
Crossref Funder Registry
ORCIDs
Presentations on Best Practices

Reproducibility



IS THERE A REPRODUCIBILITY CRISIS?

Illuminating the black box

nature

Note to biologists: submissions to *Nature* should contain complete descriptions of materials and reagents used.

Reporting Checklist For Life Sciences Articles

This checklist is used to ensure good reporting standards and
please read Reporting Life Sciences Research

Methodology

Friday, December 2, 2011 As of 12:00 AM New York 43° | 34°

THE WALL STREET JOURNAL. HEALTH

HEALTH INDUSTRY | DECEMBER 2, 2011

Scientists' Elusive Goal: Reproduc

n September, Bayer published a study describing how i

COMPUTER SCIENCE

Accessible Reproduc

Science

ICYFORUM

putation in research grows,
needed to expand recording,

Jill P. M

S

rect. M
tain a p
edgeab

The New York Times

Science

NYTimes: Home - Site Index - Archive - Help

Reliability

Nobel Laureate Retracts Two Paper

A Biostatistic Paper Alleges Potential Harm To Patients In Two Duke Clinical Studies

statistics journals r sensational
s. The most recent issue of the Annals of Applied Statistics is an

r p lleges that cancer patients
ned hat
arl
per
ost
f D
tipl

Human lives

The New York Times

Retracted Scientific Studies: A Growing List

, 2015

RETRACTED

新语丝Xin Yu Si (New Threads) Home Page

http://www.xys.org/

Most Visited Getting Started Latest Headlines Wolf-Haven Chicken Smoothie Howrse Motherload

Convert

Dartmouth and ... Aquaflow-work... Current Solicita... Description of ... 新语丝Xin Yu Si ...

新语丝

New Threads

【一】【二】【三】【yahoogroups】
Web Feed(RSS)：【国内】【海外】

d by PSI 留学生服务公司.

料，用手机浏览新语丝的方法

- 627名海内外知识分子关于肖传国派方舟子案的公开信
- 中国科学与学术诚信基金会呼吁为"扶助学术不端受害人专项基金"捐款
- 方舟子新书签名本邮购：国内、北美地区
- 网上购买：科学成就健康、批评中医、破解世界之谜、寻找生命的逻辑
- 网上购买：大象为什么不长毛 爱因斯坦信上帝吗？、带你走近科学、食品转基因
- 第十届"PSI－新语丝"网络文学奖参选作品（257篇）

Scientific integrity

No Cure

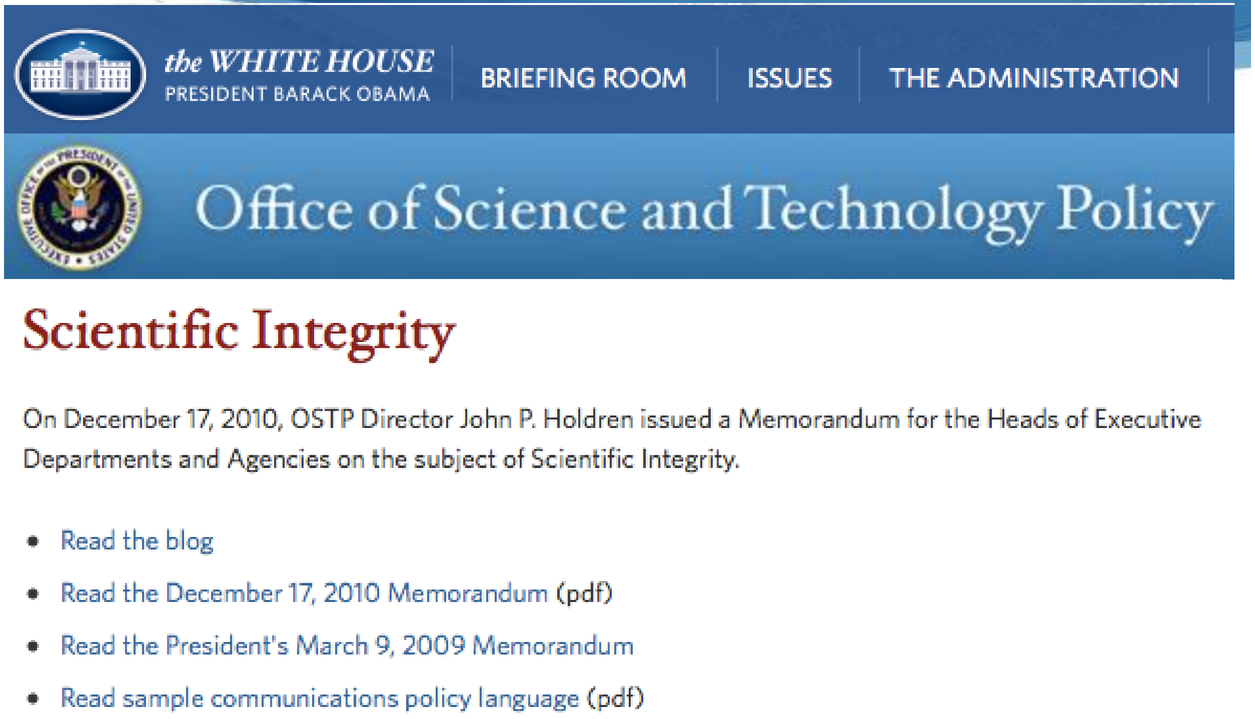
When Bayer tried to replicate results of 67 studies published in academic journals, nearly two-thirds failed.

Source: Nature Reviews Drug Discove

Financial

Trust

Government Agencies Are Changing: Scientific Integrity and Open Science

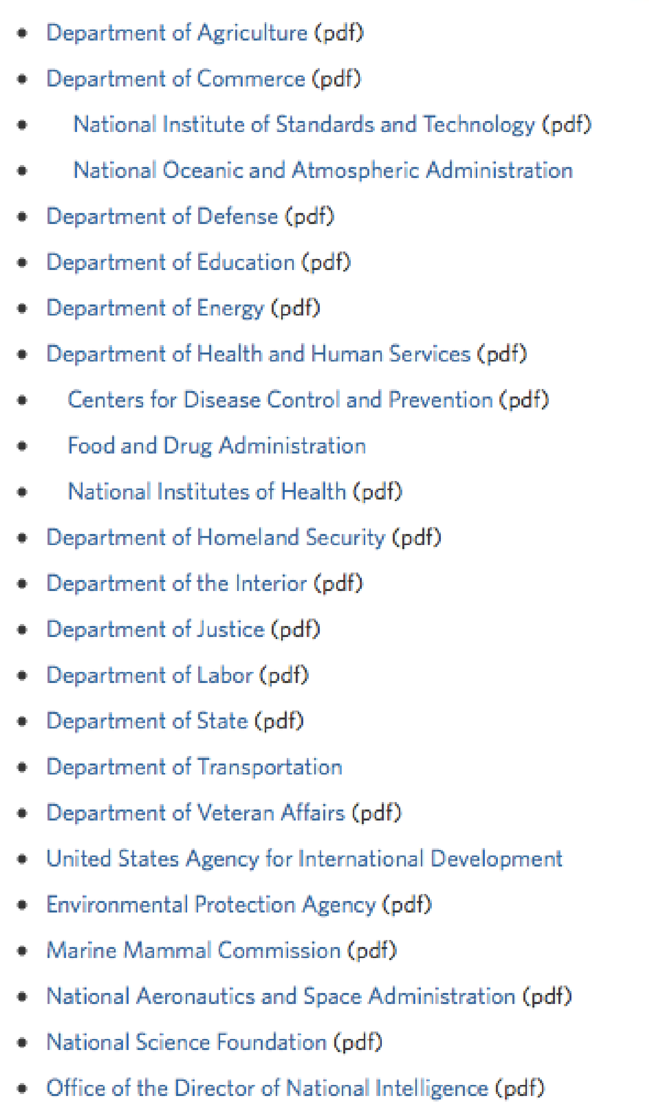


Scientific Integrity

On December 17, 2010, OSTP Director John P. Holdren issued a Memorandum for the Heads of Executive Departments and Agencies on the subject of Scientific Integrity.

- Read the blog
- Read the December 17, 2010 Memorandum (pdf)
- Read the President's March 9, 2009 Memorandum
- Read sample communications policy language (pdf)

https://obamawhitehouse.archives.gov/administration/eop/ostp/library/scientificintegrity



- Department of Agriculture (pdf)
- Department of Commerce (pdf)
- National Institute of Standards and Technology (pdf)
- National Oceanic and Atmospheric Administration
- Department of Defense (pdf)
- Department of Education (pdf)
- Department of Energy (pdf)
- Department of Health and Human Services (pdf)
- Centers for Disease Control and Prevention (pdf)
- Food and Drug Administration
- National Institutes of Health (pdf)
- Department of Homeland Security (pdf)
- Department of the Interior (pdf)
- Department of Justice (pdf)
- Department of Labor (pdf)
- Department of State (pdf)
- Department of Transportation
- Department of Veteran Affairs (pdf)
- United States Agency for International Development
- Environmental Protection Agency (pdf)
- Marine Mammal Commission (pdf)
- National Aeronautics and Space Administration (pdf)
- National Science Foundation (pdf)
- Office of the Director of National Intelligence (pdf)

Growing Importance of Scientific Integrity and Reproducibility

NSF
National Institutes of Health
DARPA
ELSEVIER
nature
AAAS
RDA
Research Data Alliance
FORCE11
The Future of Research Communications and e-Scholarship
ESIP
AGU
American Geophysical Union
CRA
Computing Research Association
The Dataverse Network Project
PLOS
figshare
cc creative commons
ICSU
CODATA
DataCite
Crossref
Altmetric
COPDESS

Science is Changing: Sharing, Open, Credit

zenodo

W3C® PROV

figshare

The Dataverse Project

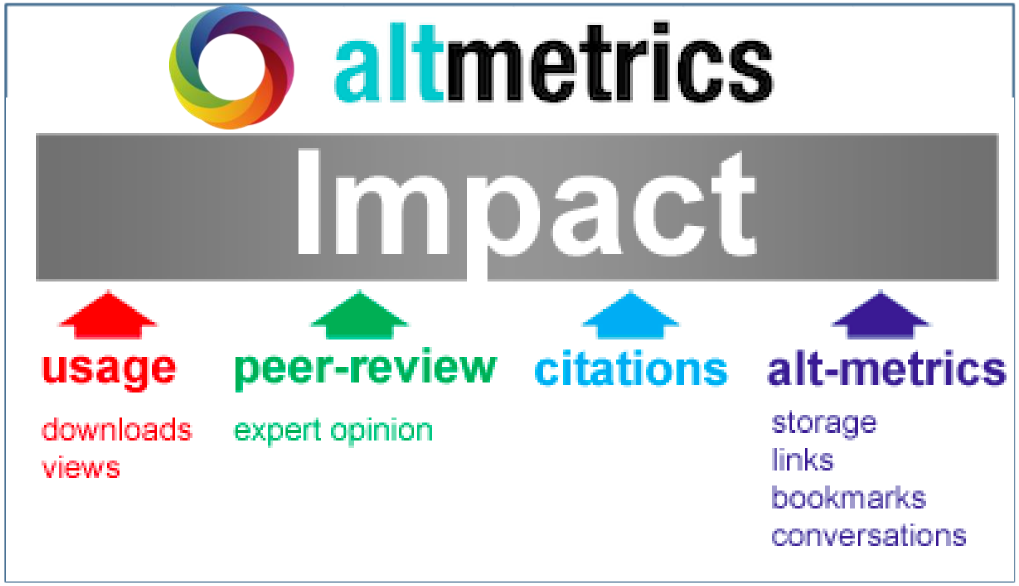
RDA RESEARCH DATA ALLIANCE

DRYAD

ESIP

PANGAEA

PLOS



altmetrics

Impact

usage
downloads
views

peer-review
expert opinion

citations

alt-metrics
storage
links
bookmarks
conversations

Impactstory

depsy

Universities are Changing: Major Initiatives in Data Science

MIT HAYSTACK OBSERVATORY

Carnegie Mellon University

ML MACHINE LEARNING DEPARTMENT

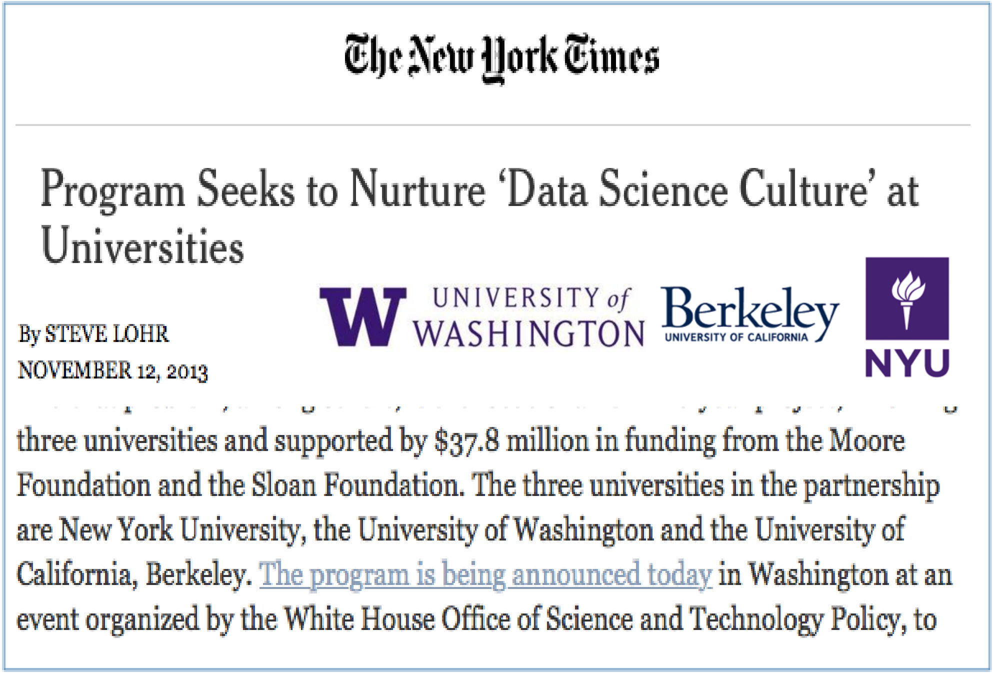
Data Science Institute COLUMBIA UNIVERSITY

NORTHEAST BIG DATA INNOVATION HUB

WEST BIG DATA INNOVATION HUB

JOHNS HOPKINS INSTITUTE FOR DATA-INTENSIVE ENGINEERING & SCIENCE

M | MIDAS MICHIGAN INSTITUTE FOR DATA SCIENCE UNIVERSITY OF MICHIGAN



The New York Times

Program Seeks to Nurture 'Data Science Culture' at Universities

By STEVE LOHR

NOVEMBER 12, 2013

UNIVERSITY of WASHINGTON

Berkeley UNIVERSITY OF CALIFORNIA

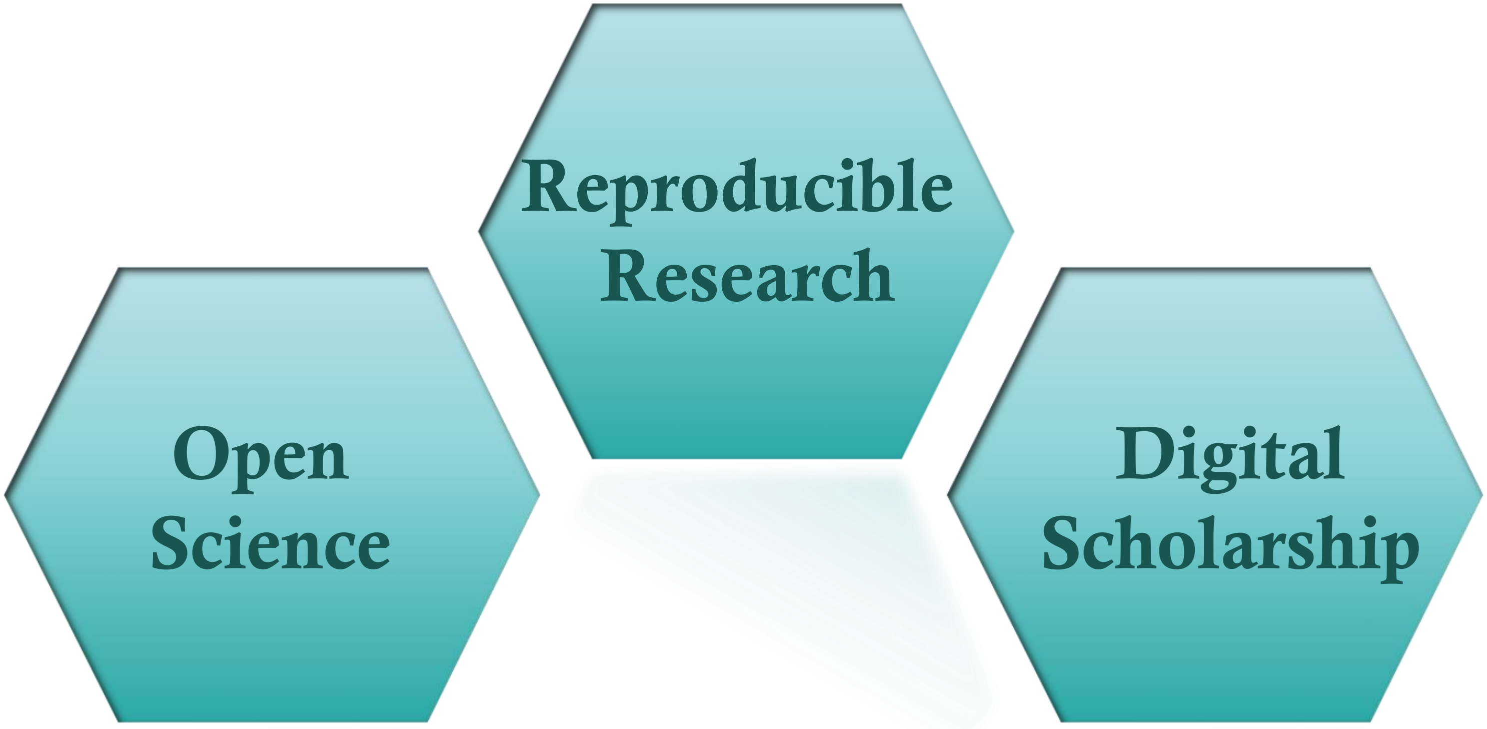
NYU

three universities and supported by $37.8 million in funding from the Moore Foundation and the Sloan Foundation. The three universities in the partnership are New York University, the University of Washington and the University of California, Berkeley. The program is being announced today in Washington at an event organized by the White House Office of Science and Technology Policy, to

CENTER FOR DATA-DRIVEN DISCOVERY Caltech

DSRC DATA SCIENCE RESEARCH CENTER Rensselaer

Core Recommendations for Scientific Publications



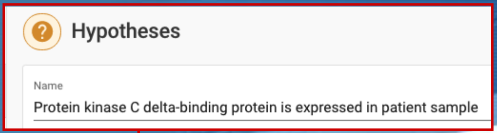
Reproducible Research

Open Science

Digital Scholarship

Reproducible Research

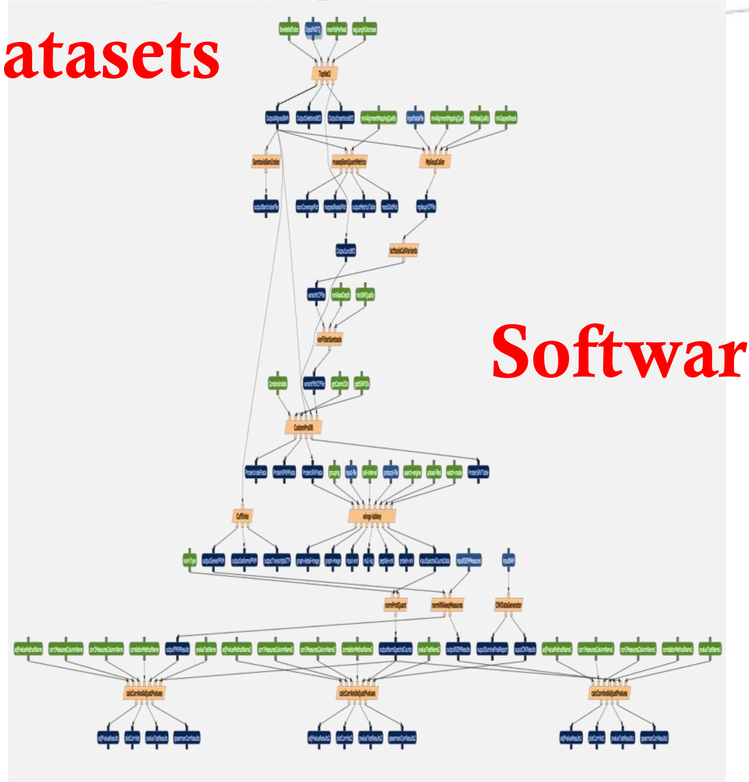
1) Reproducible Research


Hypotheses

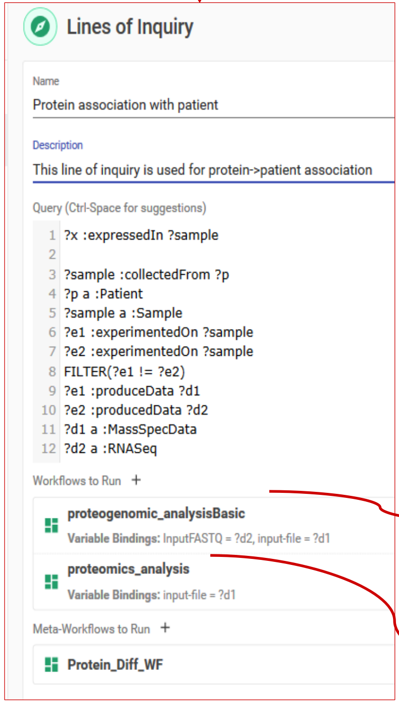
Name

Protein kinase C delta-binding protein is expressed in patient sample

Datasets



Software


Lines of Inquiry

Name

Protein association with patient

Description

This line of inquiry is used for protein->patient association

Query (Ctrl-Space for suggestions)

```
?x :expressedIn ?sample

?sample :collectedFrom ?p
?p a :Patient
?sample a :Sample
?e1 :experimentedOn ?sample
?e2 :experimentedOn ?sample
FILTER(?e1 != ?e2)
?e1 :produceData ?d1
?e2 :producedData ?d2
?d1 a :MassSpecData
?d2 a :RNASeq
```

Workflows to Run +

proteogenomic_analysisBasic

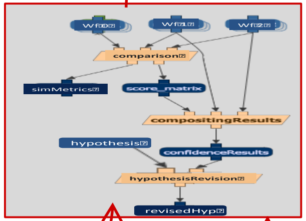
Variable Bindings: InputFASTQ = ?d2, input-file = ?d1

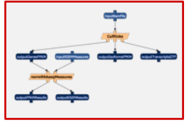
proteomics_analysis

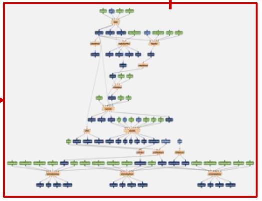
Variable Bindings: input-file = ?d1

Meta-Workflows to Run +

Protein_Diff_WF







Workflow

Experimental Design

2) Open Science

Shared repositories

zenodo

The Dataverse Project

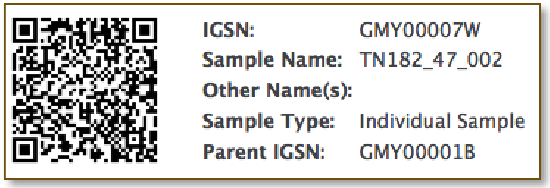
github SOCIAL CODE HOSTING

CIG COMPUTATIONAL INFRASTRUCTURE for GEODYNAMICS

Persistent unique identifiers

PURL

DOI



IGSN:	GMY00007W
Sample Name:	TN182_47_002
Other Name(s):	
Sample Type:	Individual Sample
Parent IGSN:	GMY00001B

Licenses

creative commons

CC BY

open source initiative®

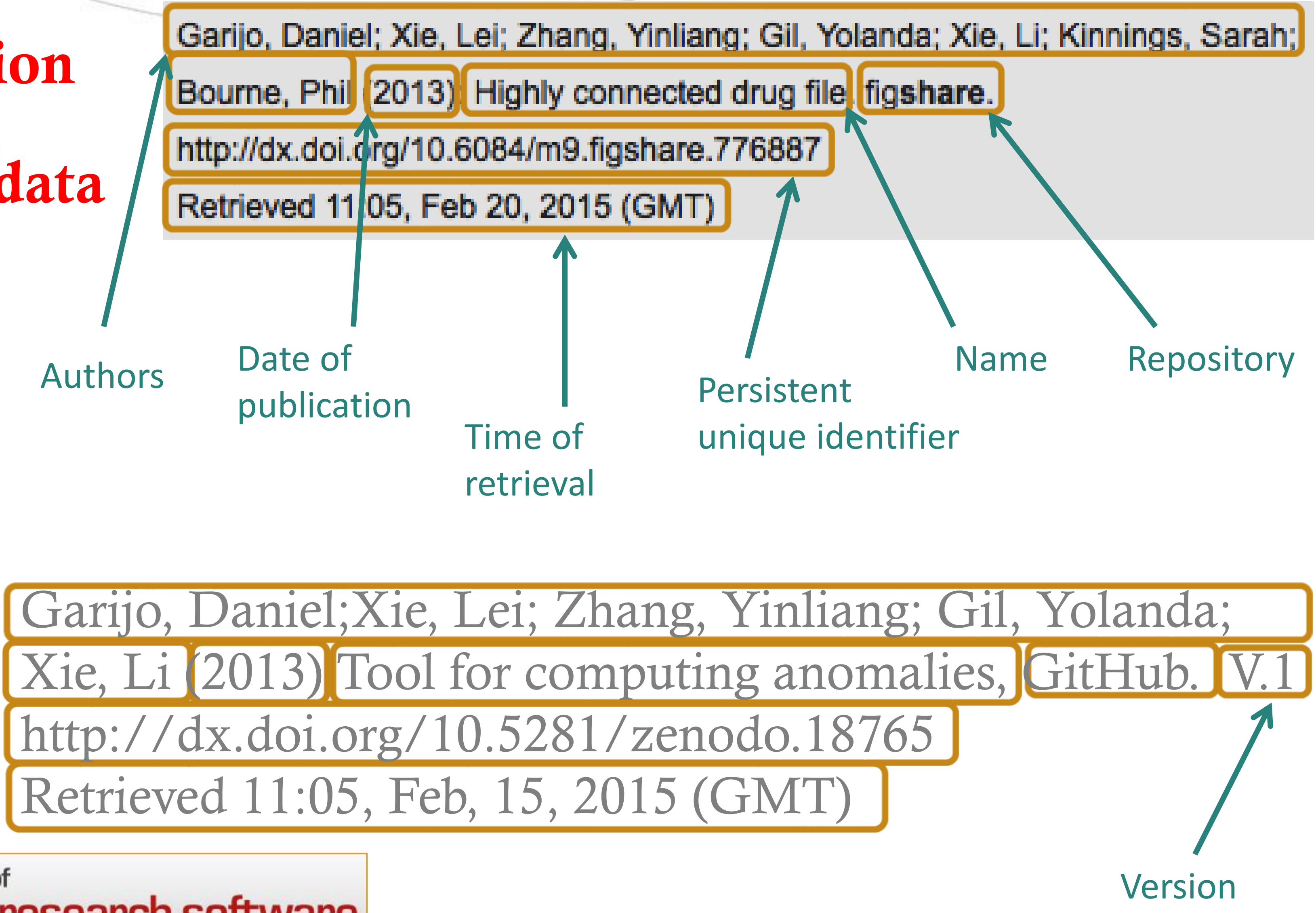
3) Digital Scholarship

Digital Scholarship

Citation

Metadata

zenodo



Garijo, Daniel; Xie, Lei; Zhang, Yinliang; Gil, Yolanda; Xie, Li; Kinnings, Sarah;
Bourne, Phil (2013) Highly connected drug file figshare.
http://dx.doi.org/10.6084/m9.figshare.776887
Retrieved 11:05, Feb 20, 2015 (GMT)

Authors

Date of publication

Time of retrieval

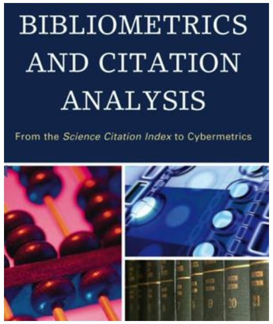
Persistent unique identifier

Name

Repository

Garijo, Daniel;Xie, Lei; Zhang, Yinliang; Gil, Yolanda;
Xie, Li (2013) Tool for computing anomalies, GitHub. V.1
http://dx.doi.org/10.5281/zenodo.18765
Retrieved 11:05, Feb, 15, 2015 (GMT)

Version

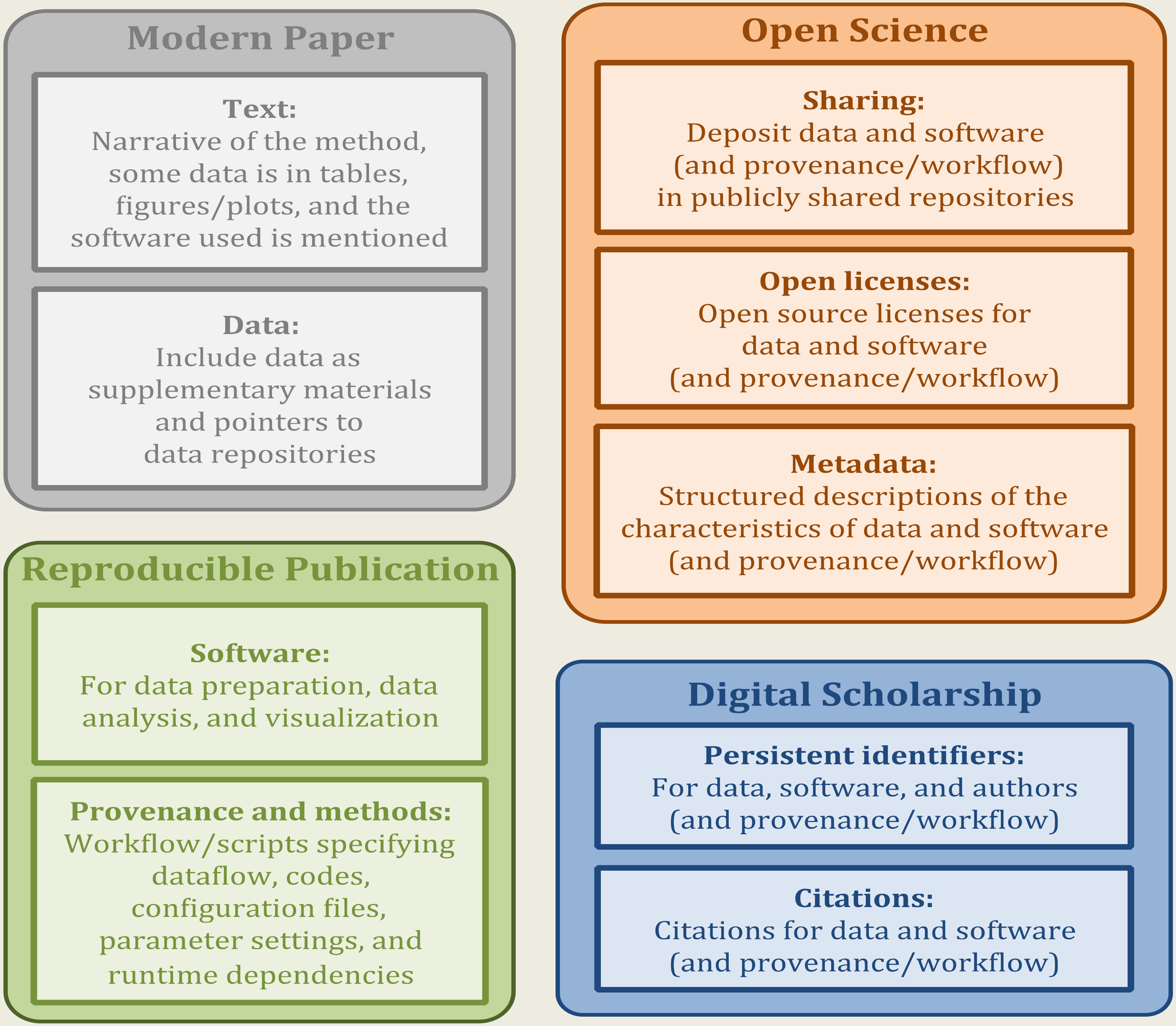


BIBLIOMETRICS AND CITATION ANALYSIS

From the *Science Citation Index* to Cybermetrics

Journal of open research software

Geoscience Paper of the Future



Modern Paper

Text:
Narrative of the method, some data is in tables, figures/plots, and the software used is mentioned

Data:
Include data as supplementary materials and pointers to data repositories

Open Science

Sharing:
Deposit data and software (and provenance/workflow) in publicly shared repositories

Open licenses:
Open source licenses for data and software (and provenance/workflow)

Metadata:
Structured descriptions of the characteristics of data and software (and provenance/workflow)

Reproducible Publication

Software:
For data preparation, data analysis, and visualization

Provenance and methods:
Workflow/scripts specifying dataflow, codes, configuration files, parameter settings, and runtime dependencies

Digital Scholarship

Persistent identifiers:
For data, software, and authors (and provenance/workflow)

Citations:
Citations for data and software (and provenance/workflow)

GEOSCIENCE PAPERS OF THE FUTURE
Onto Soft

The Geoscience Papers of the Future (GPF) Initiative

http://www.scientificpaperofthefuture.org/gpf

1. A Special Issue of a journal in all geoscience areas that includes only geoscience papers of the future

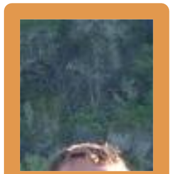
AGU American Geophysical Union

Earth and Space Science
AN OPEN ACCESS AGU JOURNAL

Special Section: Geoscience Papers of the Future

1. Training sessions for geoscientists to learn best practices in software and data sharing, provenance documentation, and scholarly publication

GPF Pioneer Authors



Cedric David, NASA/JPL
Hydrology modeling

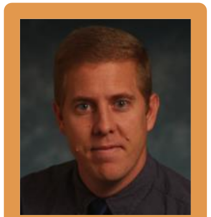
Ibrahim Demir, U. of Iowa
Hydrology sensor networks

R. W. Fulweiler, Boston U.
Biogeochemistry in marine ecology

J. Goodall/B. Essawy, U. Virginia, Hydrology/visualization

Leif Karlstrom, U. Oregon
Volcanic vent clustering

Kyo Lee, NASA/JPL
Regional climate modeling



Heith Mills, U. Houston
Geochemistry, marine biology

Ji-Hyun Oh, USC
Tropical meteorology

Suzanne Pierce, UT Austin
Hydrogeology for decision support

Allen Pope, U. Colorado
Glaciology

Mimi Tzeng, Dauphin Island Sea Lab, Ocean fisheries

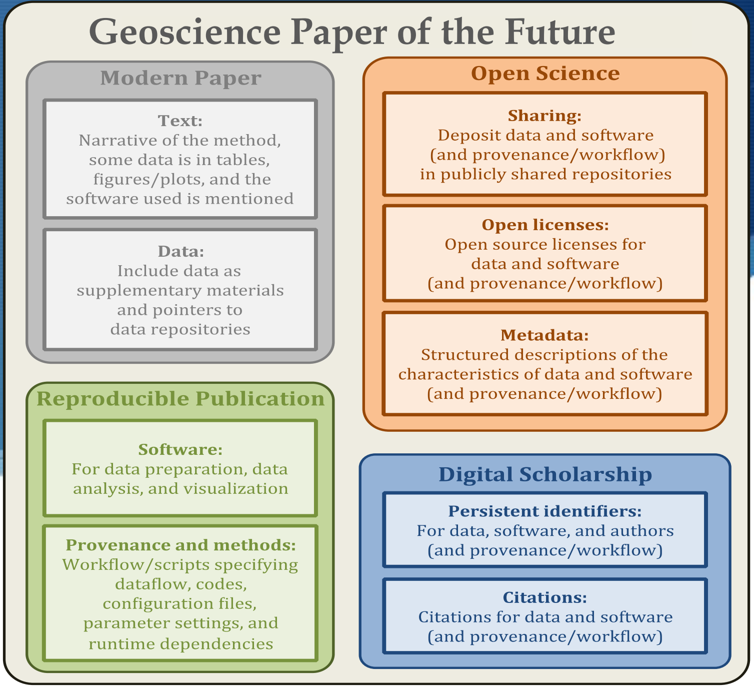
Sandra Villamizar, UC Merced
River ecohydrology

Xuan Yu, U. Delaware
Hydrologic modeling

Published Articles

www.scientificpaperofthefuture.org/gpf/special-issue

"Towards the Geoscience Paper of the Future: Best Practices for Documenting and Sharing Research from Data to Software to Provenance" Gil et al, Earth and Space Science, 2016. http://dx.doi.org/10.1002/2015EA000136



Geoscience Paper of the Future

Modern Paper

Text:
Narrative of the method, some data is in tables, figures/plots, and the software used is mentioned

Data:
Include data as supplementary materials and pointers to data repositories

Reproducible Publication

Software:
For data preparation, data analysis, and visualization

Provenance and methods:
Workflow/scripts specifying dataflow, codes, configuration files, parameter settings, and runtime dependencies

Open Science

Sharing:
Deposit data and software (and provenance/workflow) in publicly shared repositories

Open licenses:
Open source licenses for data and software (and provenance/workflow)

Metadata:
Structured descriptions of the characteristics of data and software (and provenance/workflow)

Digital Scholarship

Persistent identifiers:
For data, software, and authors (and provenance/workflow)

Citations:
Citations for data and software (and provenance/workflow)

- [David et al 2015]: 10 years of hydrology model software
- [Yu et al 2015]: Model coupling for surface/subsurface flow
- [Essawy et al 2015]: Hydrology workflows for reproducibility
- [Pope et al 2015]: Estimate subglacial lake depth from imagery
- [Fulweiler et al 2016]: Long-term estuary data & products
- [Tzeng et al 2016]: Data processing for ocean observatory
- [Demir et al 2017]: Sensor network for flood monitoring
- <*more in process*>

An Example

Understanding kinematic data from the Hellerman thrust zone

Jade Silverstein

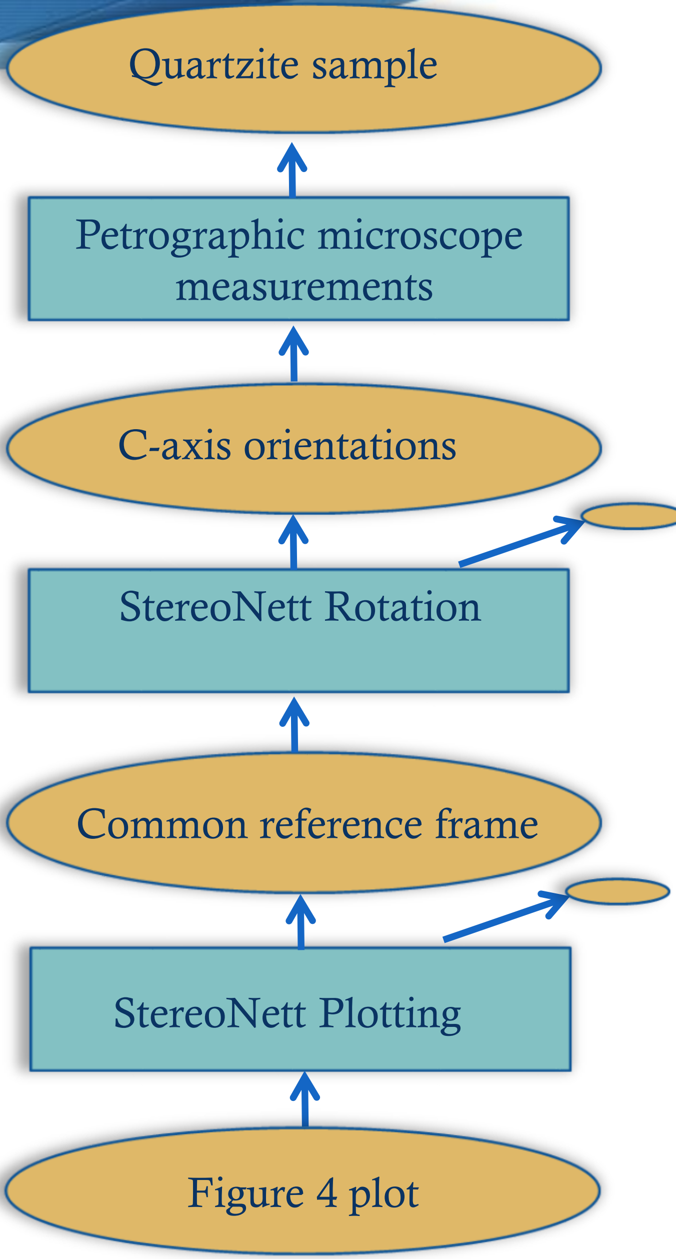
[…] We took a quartzite sample from the Hellerman thrust zone, and cut 3 thin sections. We measured c-axis orientations using a petrographic microscope. We rotated to a common reference frame using Duyster's StereoNett program. We plotted the data on lower hemisphere, equal area projections using Duyster's StereoNett program, shown in Figure 4. […]

An Example

Understanding kinematic data from the Hellerman thrust zone

Jade Silverstein

[…] We took a quartzite sample from the Hellerman thrust zone, and cut 3 thin sections. We measured c-axis orientations using a petrographic microscope. We rotated to a common reference frame using Duyster's StereoNett program. We plotted the data on lower hemisphere, equal area projections using Duyster's StereoNett program, shown in Figure 4. […]

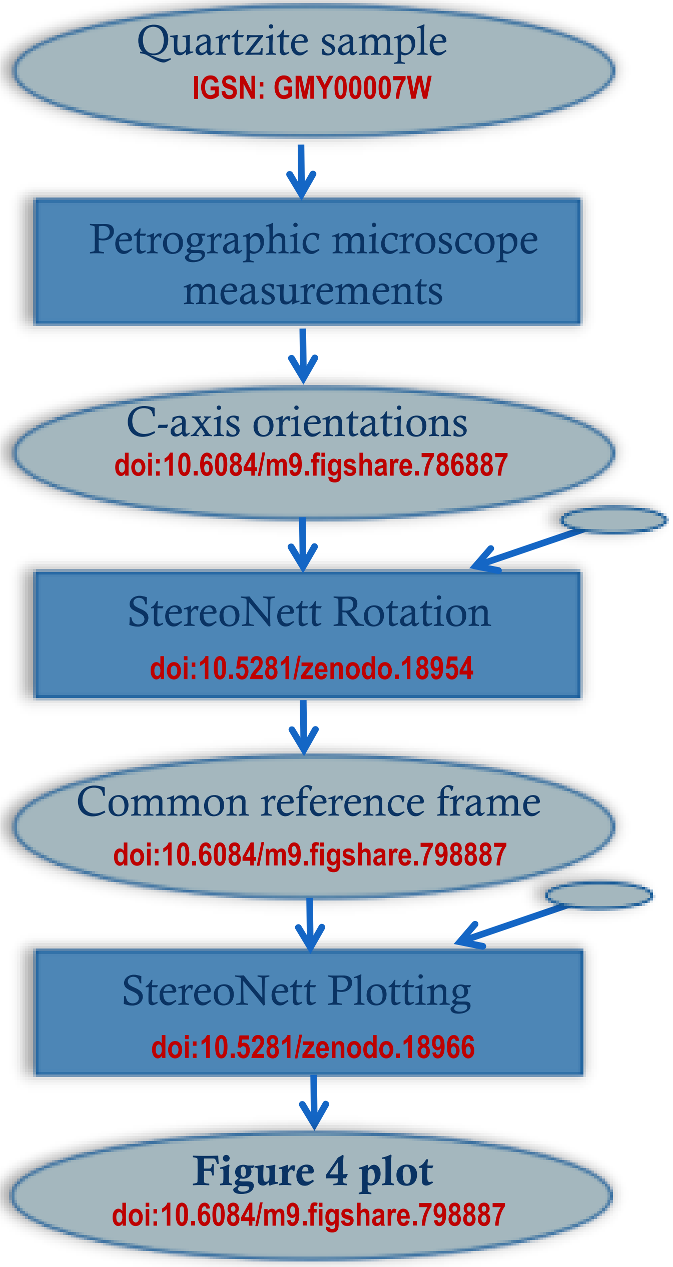


An Example

Understanding kinematic data from the Heller thrust zone (doi:10.1016/j.ess.2009.08.012)

Jade Silverstein (orcid.org/0000-0001-8455-8431)

[…] We took a quartzite sample (IGSN: GMY00007W) from the Heller thrust zone, and cut 3 thin sections. We measured c-axis orientations (doi:10.6084/m9.figshare.786887) using a petrographic microscope. We rotated to a common reference frame (doi:10.6084/m9.figshare.798887) using Duyster's StereoNett program (doi:10.5281/zenodo.18954). We plotted the data on lower hemisphere, equal area projections (doi:10.6084/m9.figshare.798887) using Duyster's StereoNett program (doi:10.5281/zenodo.18966), shown in Figure 4. The provenance is shown in Fig 5. […]



Quartzite sample
IGSN: GMY00007W

Petrographic microscope measurements

C-axis orientations
doi:10.6084/m9.figshare.786887

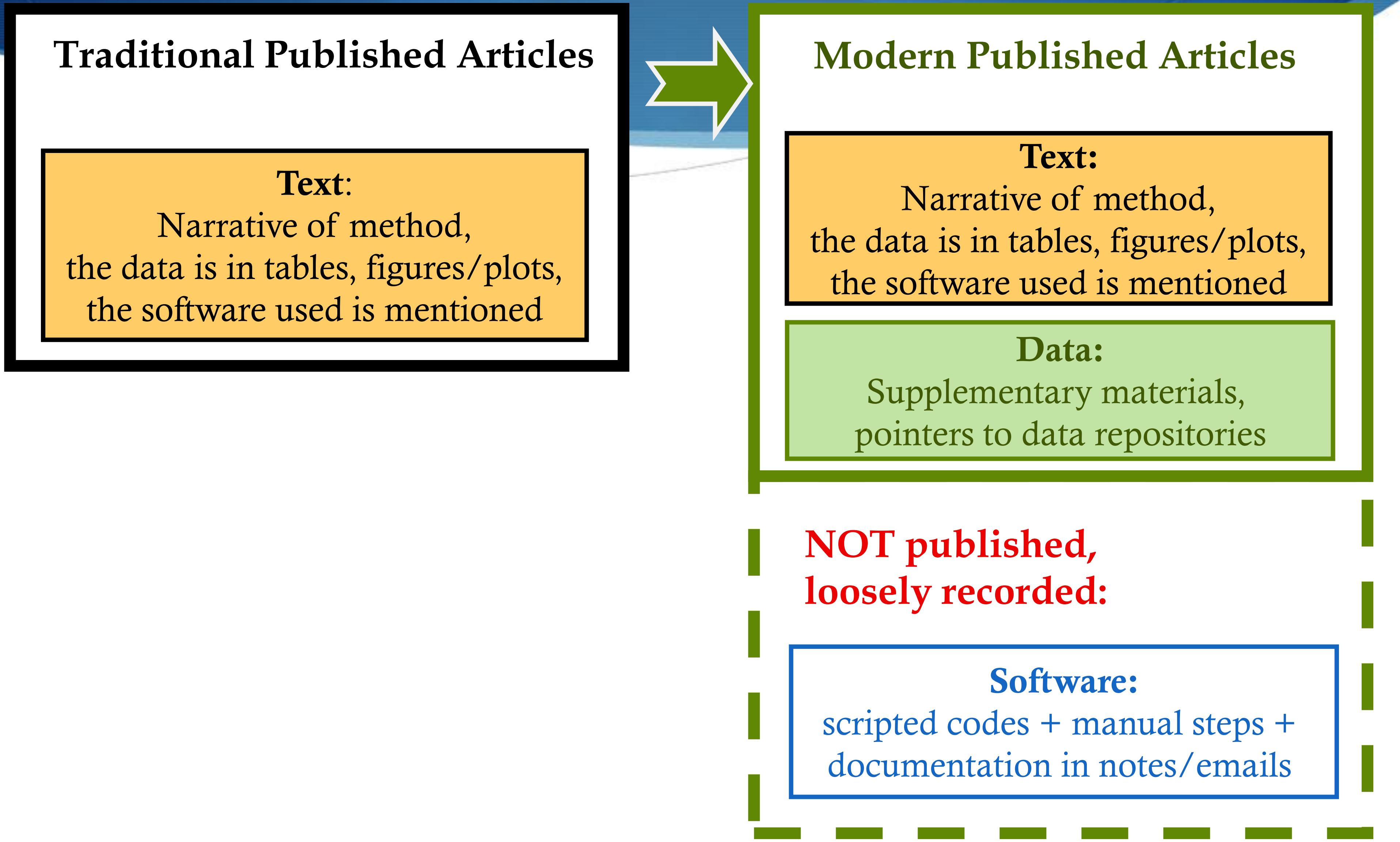
StereoNett Rotation
doi:10.5281/zenodo.18954

Common reference frame
doi:10.6084/m9.figshare.798887

StereoNett Plotting
doi:10.5281/zenodo.18966

Figure 4 plot
doi:10.6084/m9.figshare.798887

Modern Scientific Articles



Traditional Published Articles

Text:
Narrative of method,
the data is in tables, figures/plots,
the software used is mentioned

Modern Published Articles

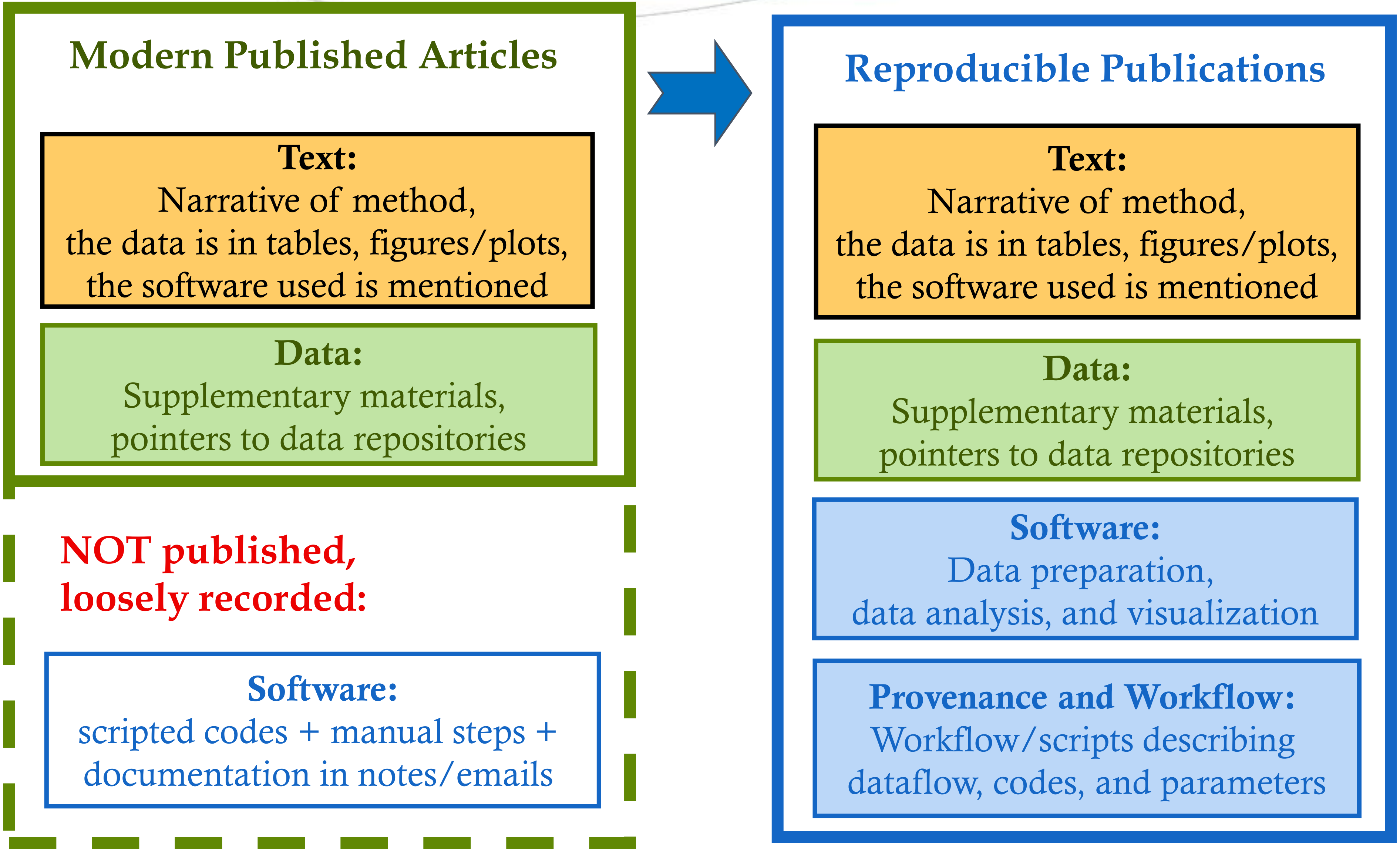
Text:
Narrative of method,
the data is in tables, figures/plots,
the software used is mentioned

Data:
Supplementary materials,
pointers to data repositories

NOT published,
loosely recorded:

Software:
scripted codes + manual steps +
documentation in notes/emails

Reproducible Articles



Modern Published Articles

Text:
Narrative of method,
the data is in tables, figures/plots,
the software used is mentioned

Data:
Supplementary materials,
pointers to data repositories

**NOT published,
loosely recorded:**

Software:
scripted codes + manual steps +
documentation in notes/emails

Reproducible Publications

Text:
Narrative of method,
the data is in tables, figures/plots,
the software used is mentioned

Data:
Supplementary materials,
pointers to data repositories

Software:
Data preparation,
data analysis, and visualization

Provenance and Workflow:
Workflow/scripts describing
dataflow, codes, and parameters

Reproducible Publications and Executable Papers

computing in SCIENCE & ENGINEERING
p. 62 Interactive Holography | p. 75 Software Development with pMatlab
REPRODUCIBLE RESEARCH
IEEE
January/February 2009, Vol. 11, No. 1

Sweave = R · LaTeX

IP[y]: Notebook

Beyond Reproducible Publications

Reproducible Publications

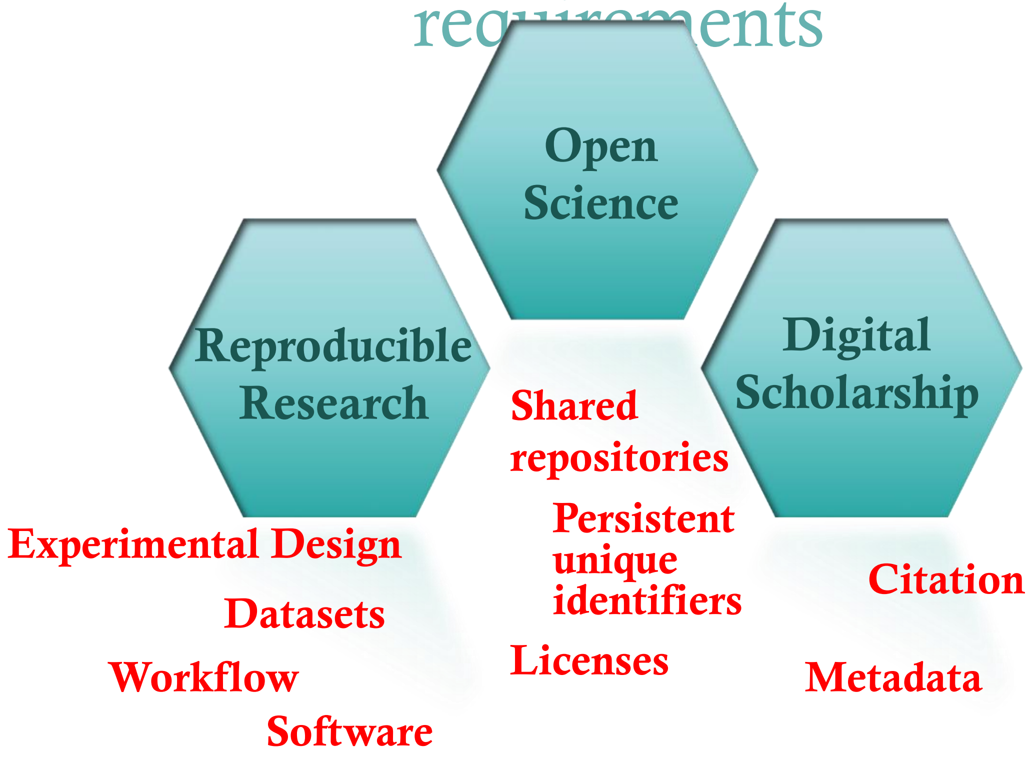
Text:
Narrative of method,
the data is in tables, figures/plots,
the software used is mentioned

Data:
Supplementary materials,
pointers to data repositories

Software:
Data preparation,
data analysis, and visualization

Provenance and methods:
Workflow/scripts describing
dataflow, codes, and parameters

The Geoscience Paper of the Future has further requirements



Open Science

Reproducible Research

Digital Scholarship

Experimental Design

Datasets

Workflow

Software

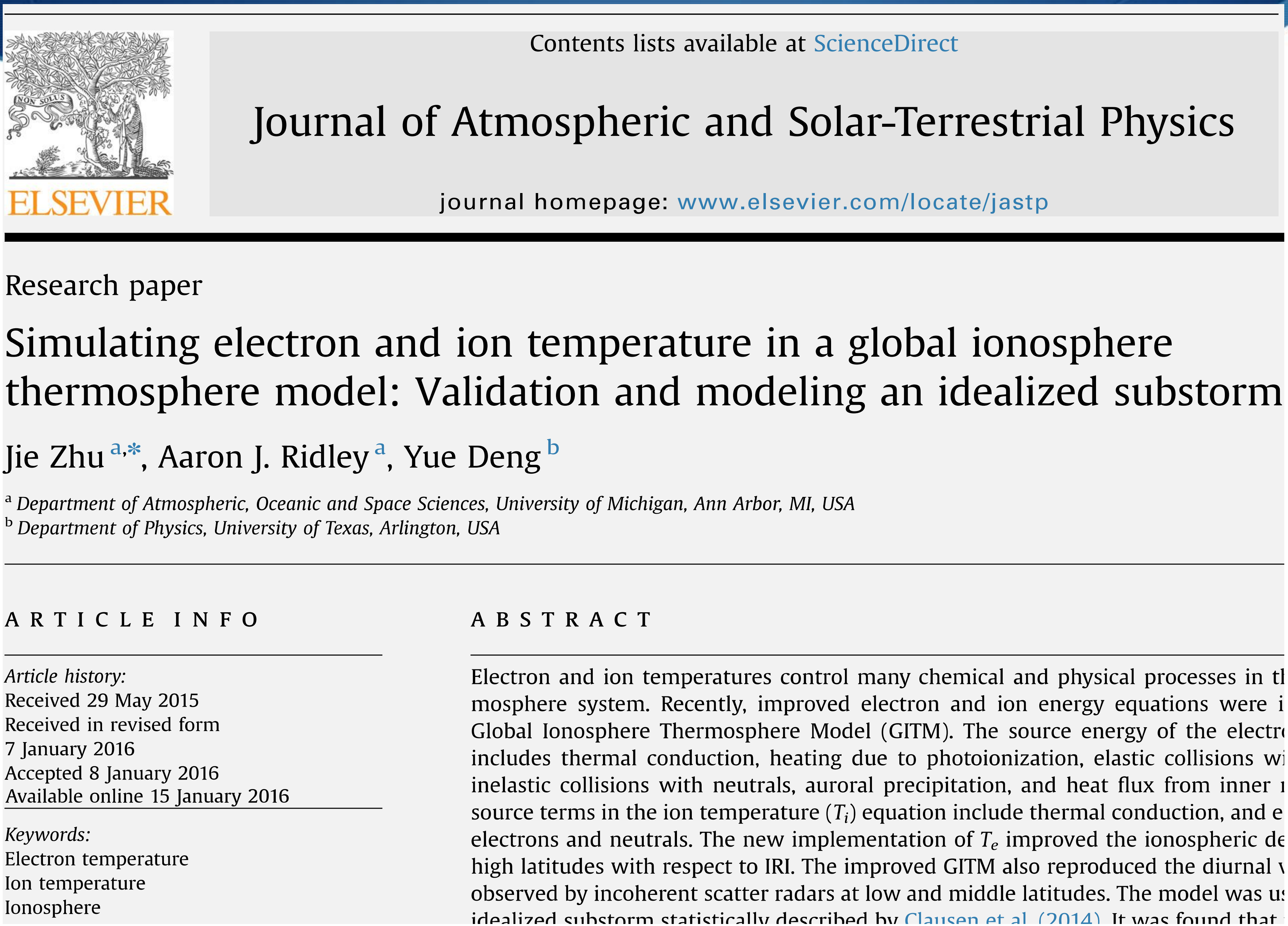
Shared repositories

Persistent unique identifiers

Licenses

Citation

Metadata



Contents lists available at ScienceDirect

Journal of Atmospheric and Solar-Terrestrial Physics

journal homepage: www.elsevier.com/locate/jastp

Research paper

Simulating electron and ion temperature in a global ionosphere thermosphere model: Validation and modeling an idealized substorm

Jie Zhu[a,*], Aaron J. Ridley[a], Yue Deng[b]

[a] Department of Atmospheric, Oceanic and Space Sciences, University of Michigan, Ann Arbor, MI, USA
[b] Department of Physics, University of Texas, Arlington, USA

ARTICLE INFO

Article history:
Received 29 May 2015
Received in revised form
7 January 2016
Accepted 8 January 2016
Available online 15 January 2016

Keywords:
Electron temperature
Ion temperature
Ionosphere

ABSTRACT

Electron and ion temperatures control many chemical and physical processes in th mosphere system. Recently, improved electron and ion energy equations were i Global Ionosphere Thermosphere Model (GITM). The source energy of the electr includes thermal conduction, heating due to photoionization, elastic collisions wi inelastic collisions with neutrals, auroral precipitation, and heat flux from inner r source terms in the ion temperature (T_i) equation include thermal conduction, and e electrons and neutrals. The new implementation of T_e improved the ionospheric de high latitudes with respect to IRI. The improved GITM also reproduced the diurnal v observed by incoherent scatter radars at low and middle latitudes. The model was us idealized substorm statistically described by Clausen et al. (2014). It was found that

Computational Aspects of the Paper

- Comparison of a new version of the Global Ionosphere Thermosphere Model (GITM) with the previous version
- Comparisons with the IRI model (Bilitza et al 2014) for several sites
- The model was used to investigate an idealized substorm

Data

From Paper

- Fig. 1 shows comparisons … using the old model (left), the new model (middle) and IRI (right) at an altitude of 400 km at 00:00 UT on December 23, 2012.
  - Missing (though provided by CCMC):
    - Brightness of sun: NOAA
    - Strength of aurora: NOAA
    - Electric fields: NASA
- Investigate the ionospheric response to an idealized substorm, which was the same as Substorm 4 investigated by Liu and Ridley (2015). The prototypical substorm was constructed based on the superposed epoch variations of IMF Bz and HP during substorms using 5-years of Challenging Minisatellite Payload (CHAMP) (Reigber et al., 2002) satellite data (Clausen et al., 2014).
  - The (Liu and Ridley 2015) paper has the substorm data and plots for it

Best Practices

- **All input data should be in a public repository (as well as any important intermediate data)**
  - **Community repositories (Madrigal), university (Dataverse), other (zenodo)…**
- **A unique identifier (DOI) should be assigned to each dataset**
- **Basic metadata should be attached**
- **A license should be specified for each dataset**
  - **Creative Commons**
    - **Recommendation: CC-BY**
- **Data should be cited in-line, the citation should be included in the references**

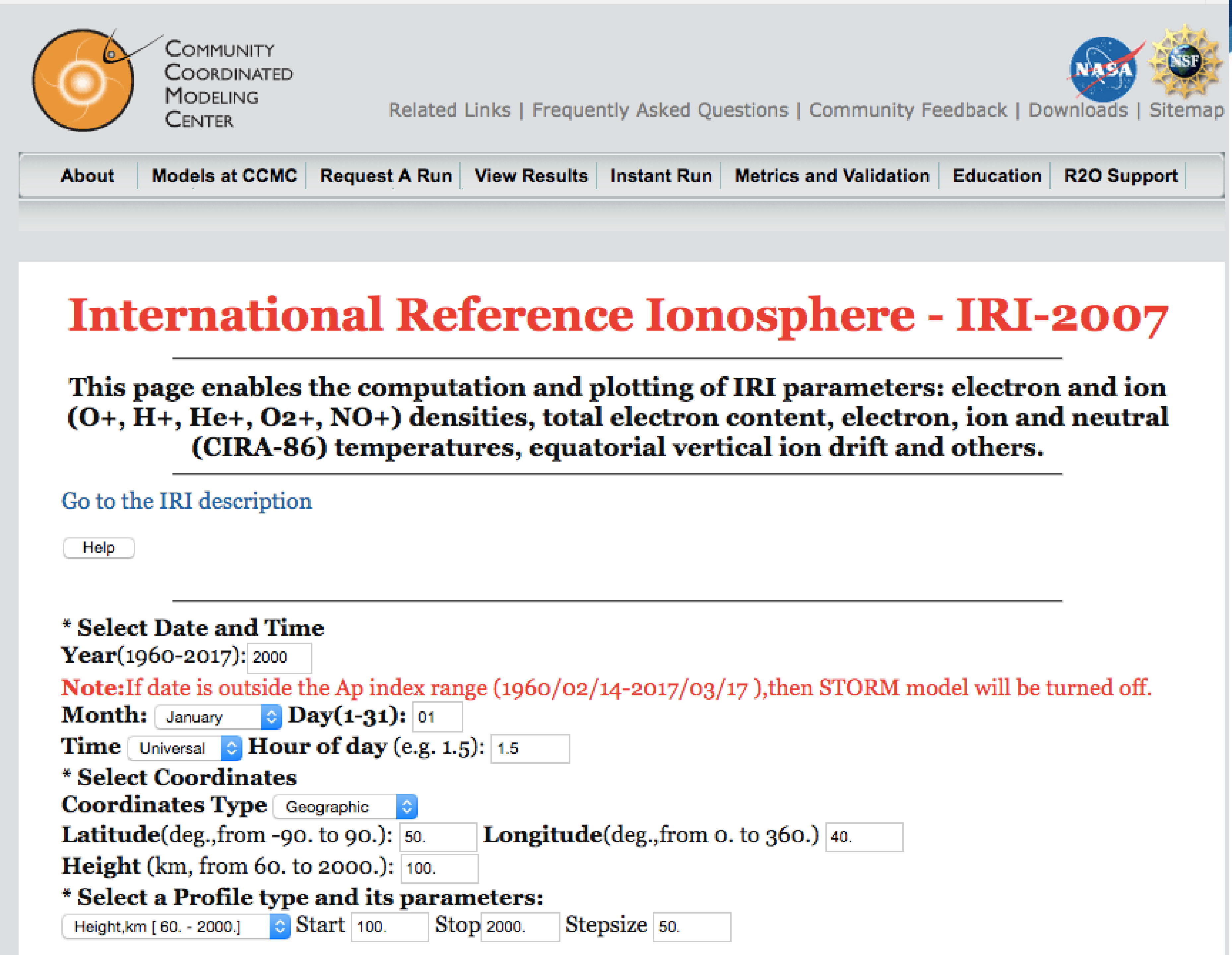
Software

From Paper

- Tables 1 and 2 present a comparison of the implementation of Te and Ti between the old and new model
  - Missing:
    - Pointers to new model and previous model software versions
- Fig. 1 shows comparisons of Ne (top), Te (middle) and Ti (bottom) using the old model (left), the new model (middle) and the International Reference Ionosphere (IRI) model (Bilitza et al., 2014) (right)
  - Missing:
    - Pointer to ISI software version

Best Practices

- **All software should be in a public repository**
  - **Community repository**
    - **GitHub, or zenodo**
- **Unique identifier (DOI) should be assigned to each software version**
- **Basic metadata should be attached**
  - **See www.ontosoft.org**
- **License should be specified**
  - **See www.opensource.org**
    - **Recommend: Apache v2, MIT**
- **Software should be cited in-line (including version), the citation should be included in the references**


Community Coordinated Modeling Center
NASA
NSF
Related Links | Frequently Asked Questions | Community Feedback | Downloads | Sitemap
About
Models at CCMC
Request A Run
View Results
Instant Run
Metrics and Validation
Education
R2O Support
Help

International Reference Ionosphere - IRI-2007

This page enables the computation and plotting of IRI parameters: electron and ion (O+, H+, He+, O2+, NO+) densities, total electron content, electron, ion and neutral (CIRA-86) temperatures, equatorial vertical ion drift and others.

Go to the IRI description

*** Select Date and Time**

Year(1960-2017): 2000

Note: If date is outside the Ap index range (1960/02/14-2017/03/17),then STORM model will be turned off.

Month: January **Day(1-31):** 01

Time Universal **Hour of day** (e.g. 1.5): 1.5

*** Select Coordinates**

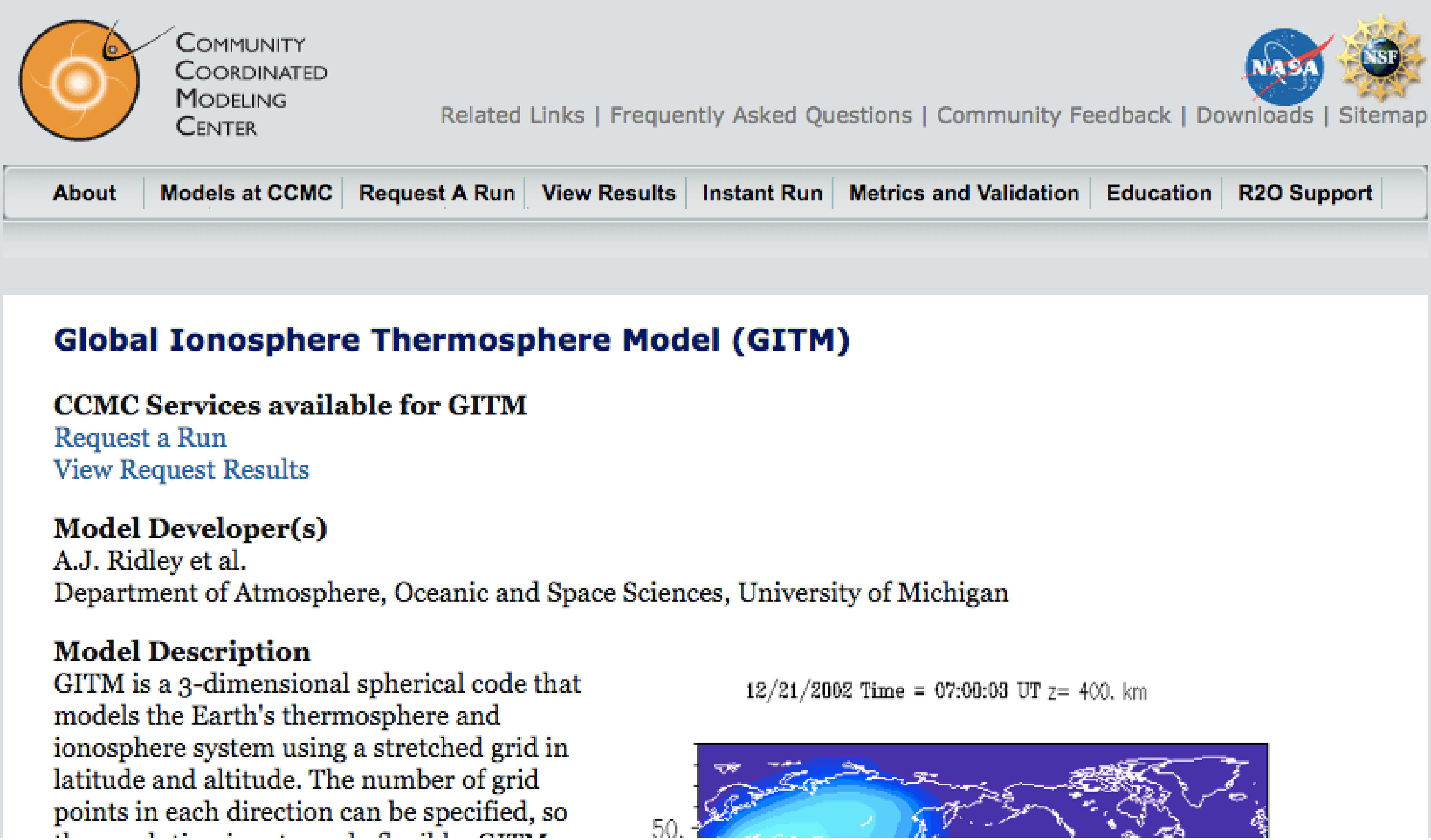
Coordinates Type Geographic

Latitude(deg.,from -90. to 90.): 50. **Longitude**(deg.,from 0. to 360.) 40.

Height (km, from 60. to 2000.): 100.

*** Select a Profile type and its parameters:**

Height,km [60. - 2000.] Start 100. Stop 2000. Stepsize 50.


Community Coordinated Modeling Center
Related Links | Frequently Asked Questions | Community Feedback | Downloads | Sitemap
About
Models at CCMC
Request A Run
View Results
Instant Run
Metrics and Validation
Education
R2O Support
12/21/2002 Time = 07:00:03 UT z= 400. km
50.

Global Ionosphere Thermosphere Model (GITM)

CCMC Services available for GITM
Request a Run
View Request Results

Model Developer(s)
A.J. Ridley et al.
Department of Atmosphere, Oceanic and Space Sciences, University of Michigan

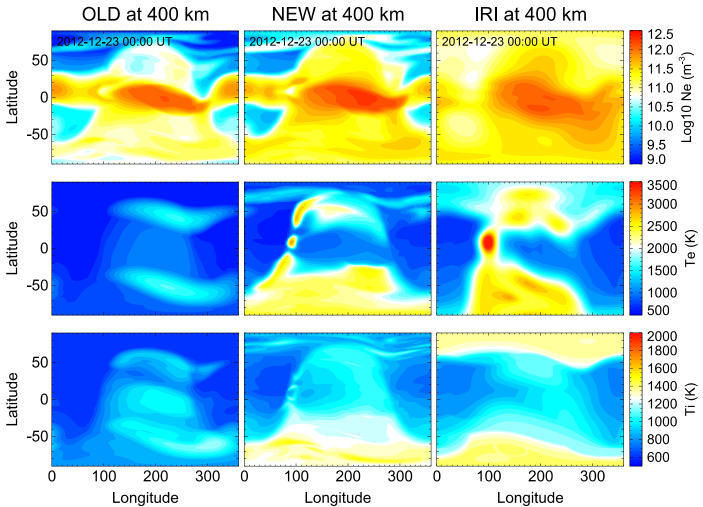
Model Description
GITM is a 3-dimensional spherical code that models the Earth's thermosphere and ionosphere system using a stretched grid in latitude and altitude. The number of grid points in each direction can be specified, so

- ★ Accessing older versions?
- ★ Identifying older versions uniquely?

Workflow

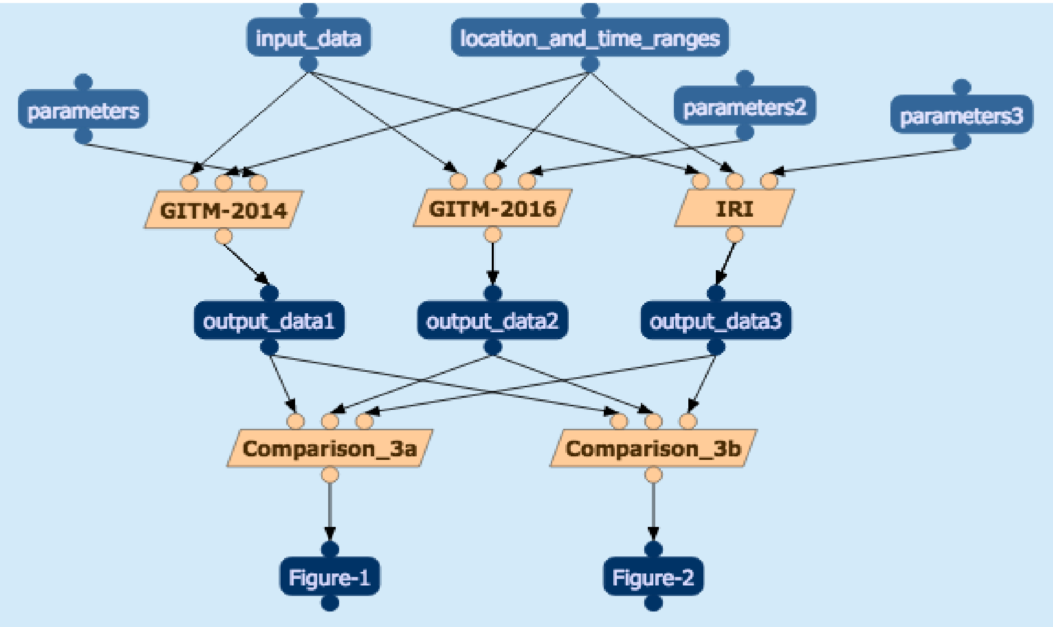
From Paper

- ★ Fig. 1 shows comparisons of Ne (top), Te (middle) and Ti (bottom) using the old model (left), the new model (middle) and the International Reference Ionosphere (IRI) model (Bilitza et al., 2014) (right)



Best Practices

- ★ **Sketch a workflow diagram**
  - ★ **Software invocations**
  - ★ **Dataflow connections**
- ★ **Include workflow diagram in figure or supplementary materials**



Provenance of Results

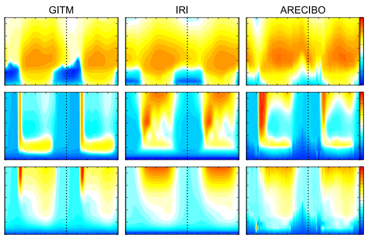
From Paper

Fig. 5 shows a comparison between GITM, IRI and measurement by the Arecibo ISR from 100 km to 650 km on April 13th and 14th, 2013. The missing data in the observations were filled by an altitudinal linear interpolation.

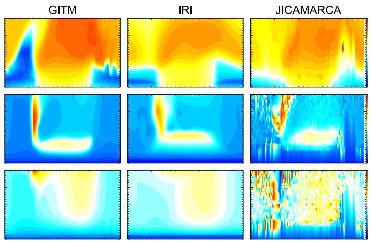
Best Practices

- ★ **Specification of all input data and parameters**
- ★ **Publication of intermediate and final results facilitates reproducibility**

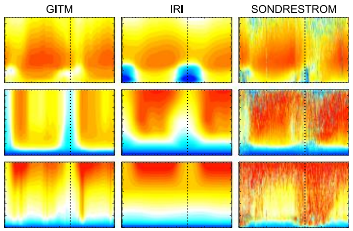
ARECIBO

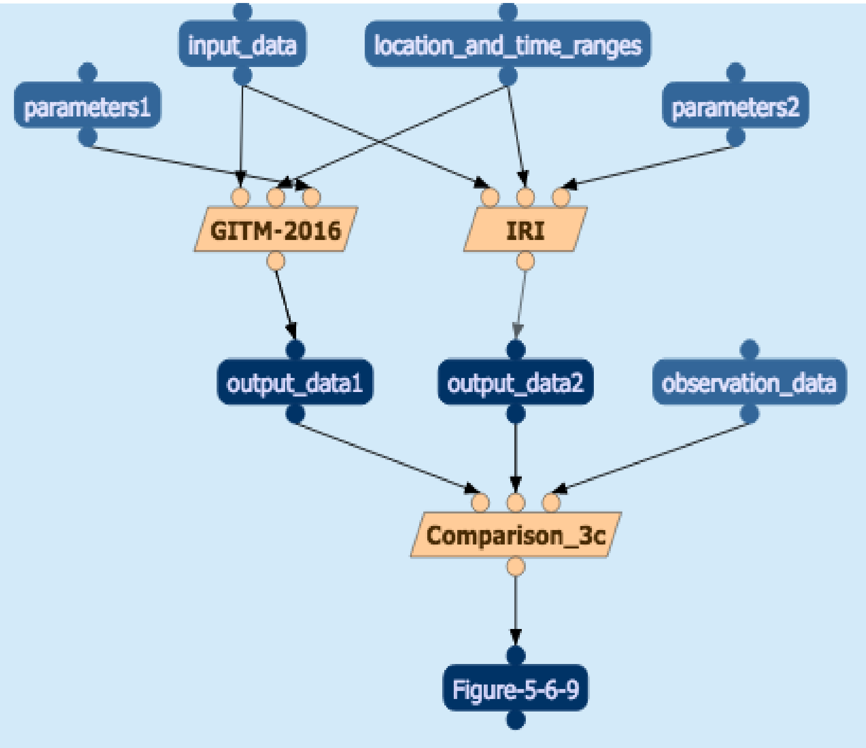


JICAMARCA



SONDRESTROM





Author Checklist

1. Data accessibility
2. Data documentation
3. Software accessibility
4. Software documentation
5. Methods documentation
6. Provenance documentation
7. Author identification

- ★ **For datasets**, the paper should include one or more citations, specifying the authors, the site where they are described and can be accessed, the repository, and the license.
- ★ **For software**, the paper should include one or more citations, specifying the authors, the site where it is described and can be accessed, the repository, and the license.
- ★ **For provenance and workflow**, the paper should include figures and traces, and if available the citations mentioning the authors, site to access them, the repository, and the license.
- ★ **For authors**, each should have a unique identifier (e.g., ORCID)

Incorporate GPF Best Practices Into Your Work

- Easier to track research products, train new lab members, build on prior work, etc.
- Making a paper into a GPF is then very straightforward

(Photo/Getty Images) (Girl brushing teeth)
http://nbclatino.com/2012/06/04/untreated-dental-cavities-soar-among-latino-kids/

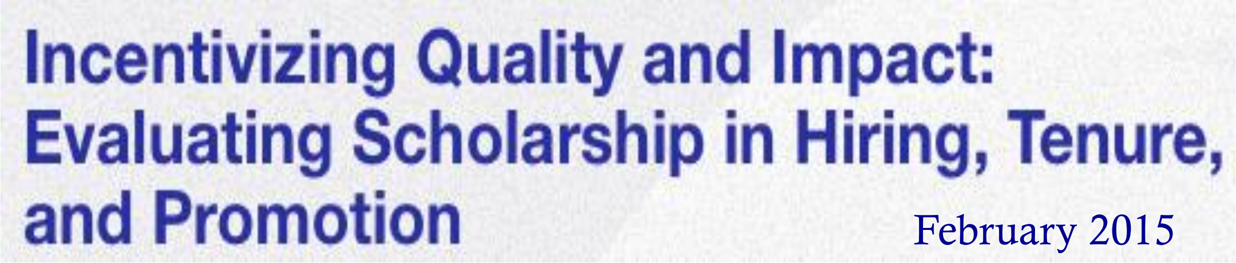
Why Learn to Write a Scientific Paper of the Future

1. Practice **open science and reproducible research**
2. **Get credit** for all your research products
   - Citations for software, data, samples, …
3. **Increase citations** of your papers
4. Write impressive **Data Management Plans**
5. **Extend your CV** with data and software sections
6. Improve the **management of your research assets**
7. **Reproduce** your work from years ago and build on it
8. Address new **funder and journal requirements**
9. Attract **transformative students**
10. Demonstrate **leadership** by stepping into the future

Recommendations from Scientific Societies

CRA

Computing Research Association



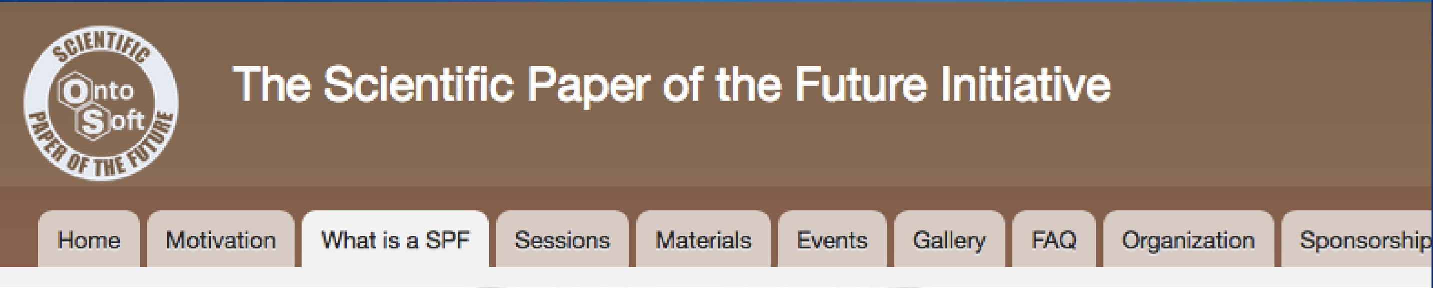
Incentivizing Quality and Impact: Evaluating Scholarship in Hiring, Tenure, and Promotion

February 2015

"The field benefits when researchers build on each other's work. To do so, requires that research advances be accompanied by discussion of methods, comparisons with related work, inclusion of supporting data and proofs, access to artifacts, and other details. Certain publication formats and review processes, however, encourage practices inconsistent with these elements of good scholarship. ***Length restrictions often are satisfied by omitting critical content, which hinders reproducing the results, understanding their novelty, or delimiting a contribution's applicability. The omission of supporting data and proofs, also common practice, hobbles efforts to validate or extend the work****."*

http://cra.org/resources/best-practice-memos/incentivizing-quality-and-impact-evaluating-scholarship-in-hiring-tenure-and-promotion/

www.scientificpaperofthefuture.org



The Scientific Paper of the Future Initiative

Home | Motivation | What is a SPF | Sessions | Materials | Events | Gallery | FAQ | Organization | Sponsorship

The Scientific Paper of the Future

http://www.scientificpaperofthefuture.org

OntoSoft Training

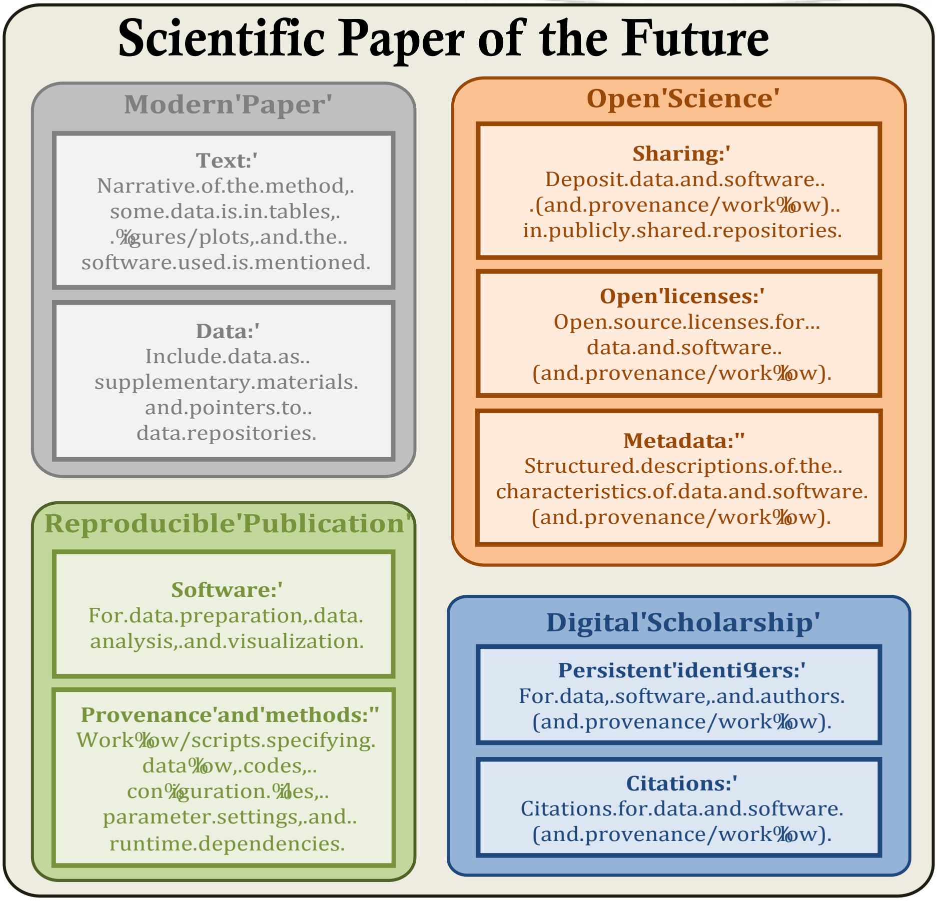
October 2016

ontosoft@gmail.com

http://dx.doi.org/10.5281/zenodo.159206

CC-BY Attribution

EarthCube



Scientific Paper of the Future

Modern Paper

Text: Narrative of the method, some data is in tables, figures/plots, and the software used is mentioned.

Data: Include data as supplementary materials and pointers to data repositories.

Reproducible Publication

Software: For data preparation, data analysis, and visualization.

Provenance and methods: Workflow/scripts specifying data flow, codes, configuration files, parameter settings, and runtime dependencies.

Open Science

Sharing: Deposit data and software (and provenance/workflow) in publicly shared repositories.

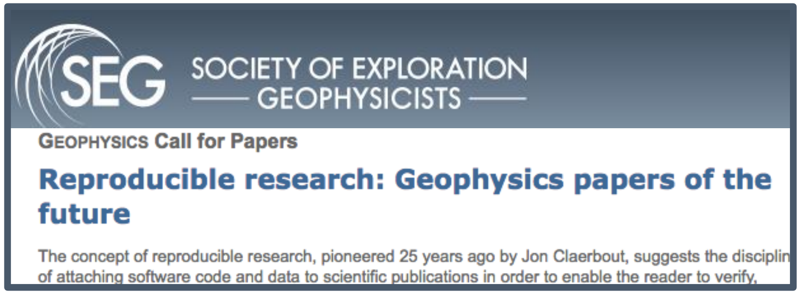
Open licenses: Open source licenses for data and software (and provenance/workflow).

Metadata: Structured descriptions of the characteristics of data and software (and provenance/workflow).

Digital Scholarship

Persistent identifiers: For data, software, and authors (and provenance/workflow).

Citations: Citations for data and software (and provenance/workflow).

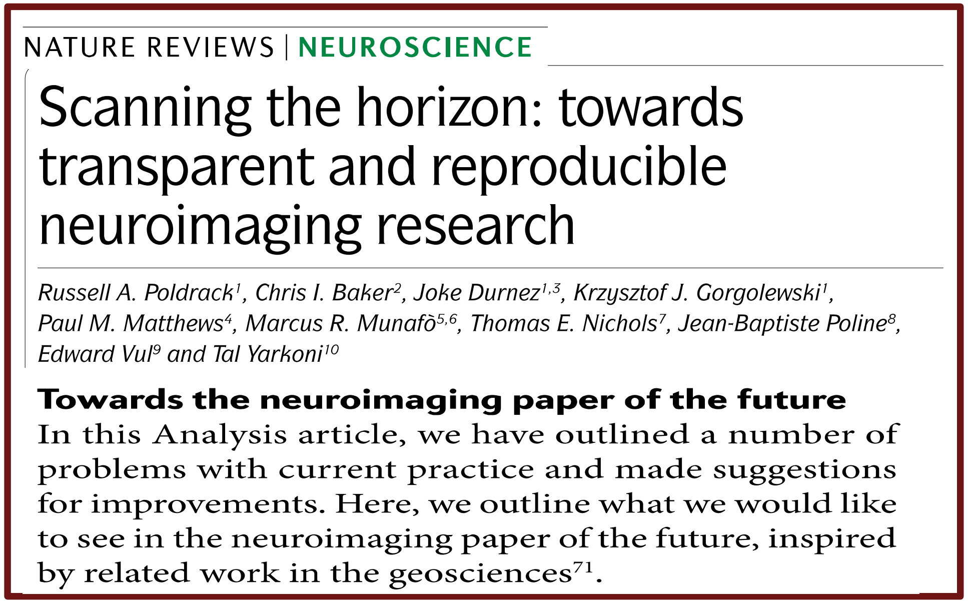


SEG SOCIETY OF EXPLORATION GEOPHYSICISTS

GEOPHYSICS Call for Papers

Reproducible research: Geophysics papers of the future

The concept of reproducible research, pioneered 25 years ago by Jon Claerbout, suggests the disciplin
of attaching software code and data to scientific publications in order to enable the reader to verify,



NATURE REVIEWS | NEUROSCIENCE

Scanning the horizon: towards transparent and reproducible neuroimaging research

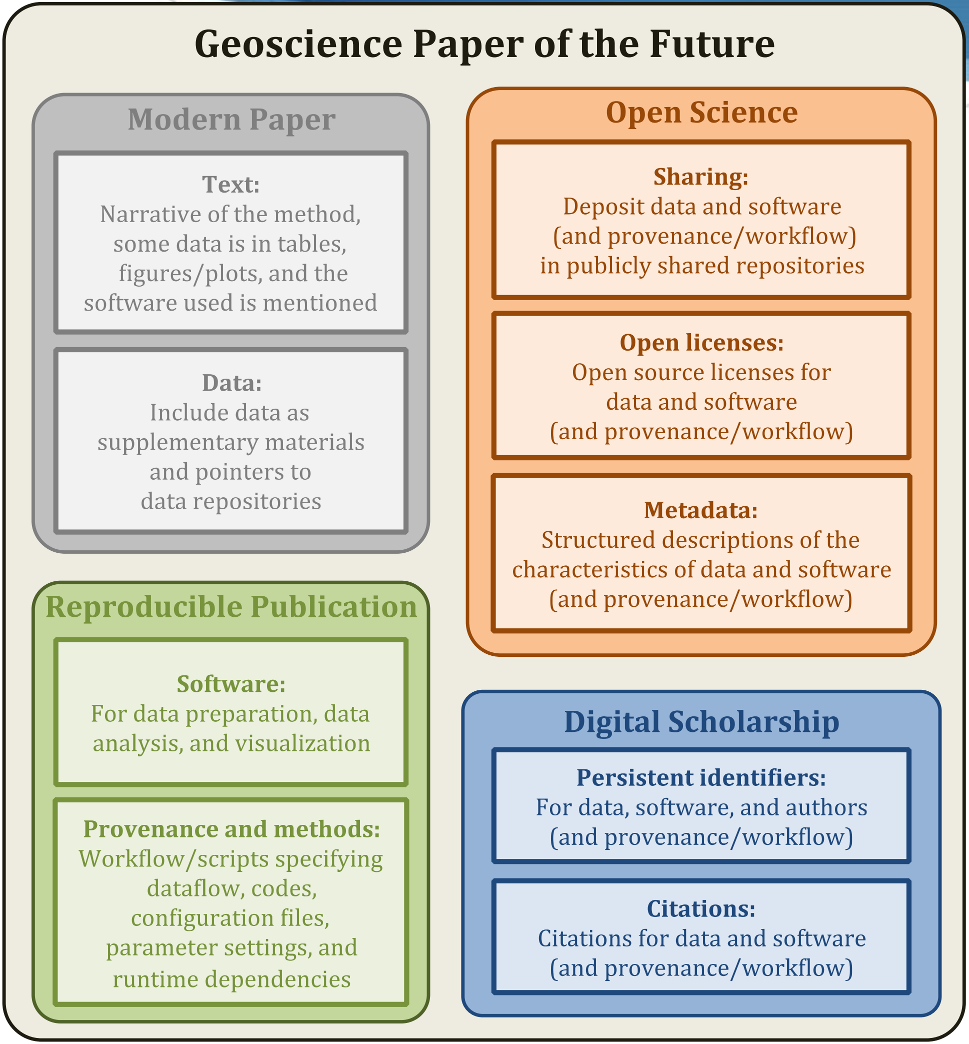
Russell A. Poldrack[1], Chris I. Baker[2], Joke Durnez[1,3], Krzysztof J. Gorgolewski[1], Paul M. Matthews[4], Marcus R. Munafò[5,6], Thomas E. Nichols[7], Jean-Baptiste Poline[8], Edward Vul[9] and Tal Yarkoni[10]

Towards the neuroimaging paper of the future

In this Analysis article, we have outlined a number of problems with current practice and made suggestions for improvements. Here, we outline what we would like to see in the neuroimaging paper of the future, inspired by related work in the geosciences[71].

For More Information

http://www.scientificpaperofthefuture.org/gpf



Geoscience Paper of the Future

Modern Paper

Text:
Narrative of the method, some data is in tables, figures/plots, and the software used is mentioned

Data:
Include data as supplementary materials and pointers to data repositories

Reproducible Publication

Software:
For data preparation, data analysis, and visualization

Provenance and methods:
Workflow/scripts specifying dataflow, codes, configuration files, parameter settings, and runtime dependencies

Open Science

Sharing:
Deposit data and software (and provenance/workflow) in publicly shared repositories

Open licenses:
Open source licenses for data and software (and provenance/workflow)

Metadata:
Structured descriptions of the characteristics of data and software (and provenance/workflow)

Digital Scholarship

Persistent identifiers:
For data, software, and authors (and provenance/workflow)

Citations:
Citations for data and software (and provenance/workflow)

GPF recommended best practices:
http://dx.doi.org/10.1002/2015EA000136

Special issue:
http://tinyurl.com/ess-gpf

Training materials:
http://dx.doi.org/10.5281/zenodo.15920

GEOSCIENCE PAPERS OF THE FUTURE

OntoSoft

InGeO Integrated Geoscience Observatory

NSF

EarthCube!

ICER-1440323
ICER-1343800

ICER-1541057

EXTRA SLIDES:

OVERVIEW OF GPF RECOMMENDATIONS AND AUTHOR CHECKLIST

Author Checklist

1. Data accessibility
2. Data documentation
3. Software accessibility
4. Software documentation
5. Methods documentation
6. Provenance documentation
7. Author identification

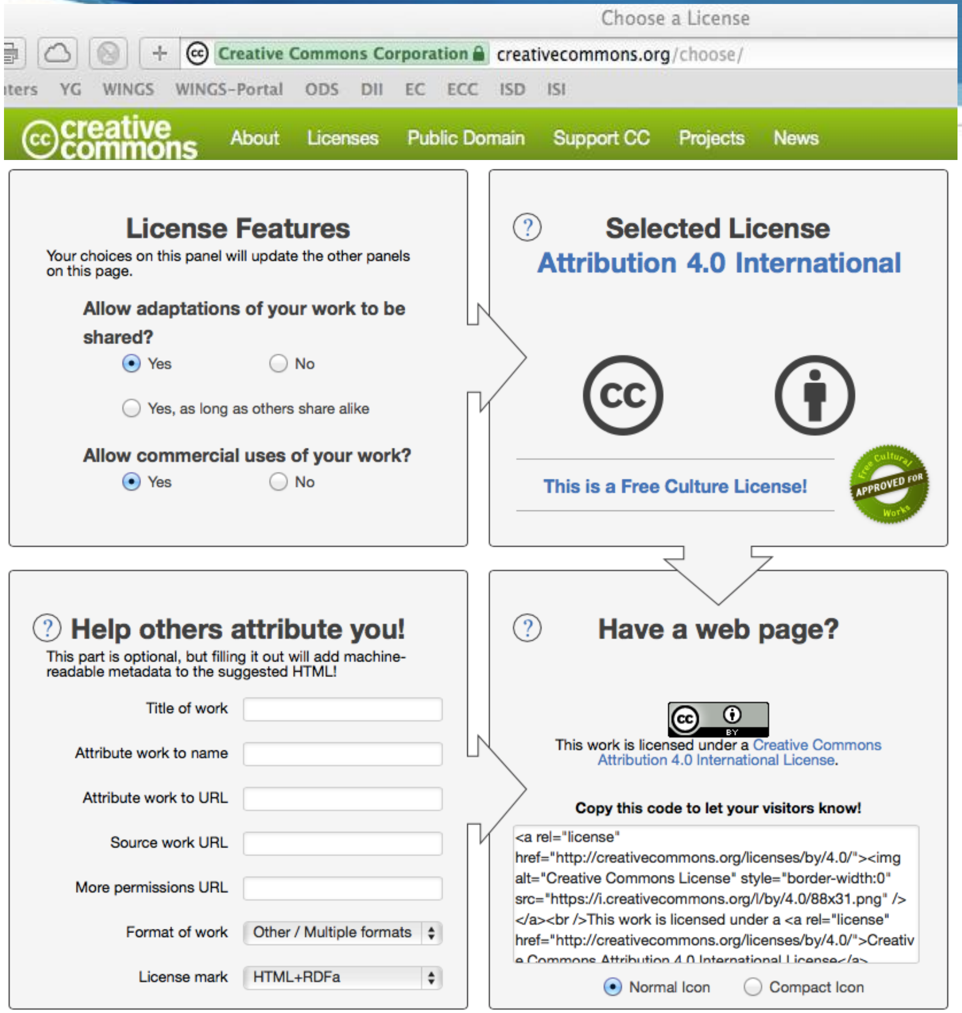
- ★ **For datasets**, the paper should include one or more citations, specifying the authors, the site where they are described and can be accessed, the repository, and the license.
- ★ **For software**, the paper should include one or more citations, specifying the authors, the site where it is described and can be accessed, the repository, and the license.
- ★ **For provenance and workflow**, the paper should include figures and traces, and if available the citations mentioning the authors, site to access them, the repository, and the license.
- ★ **For authors**, each should have a unique identifier (e.g., ORCID)

Directories of Research Data Repositories

http://www.thestaffingstream.com/2012/08/06/the-buzz-about-talent-communities/

- http://www.re3data.org
- http://databib.org/index_subjects.php
- http://oad.simmons.edu/oadwiki/Data_repositories
- http://www.force11.org
- http://www.nature.com/sdata/data-policies/repositories

Choose a License



License Features

Your choices on this panel will update the other panels on this page.

Allow adaptations of your work to be shared?

● Yes ○ No

○ Yes, as long as others share alike

Allow commercial uses of your work?

● Yes ○ No

Selected License

Attribution 4.0 International

This is a Free Culture License!

Help others attribute you!

This part is optional, but filling it out will add machine-readable metadata to the suggested HTML!

Title of work

Attribute work to name

Attribute work to URL

Source work URL

More permissions URL

Format of work: Other / Multiple formats

License mark: HTML+RDFa

Have a web page?

This work is licensed under a Creative Commons Attribution 4.0 International License.

Copy this code to let your visitors know!

```
<a rel="license"
href="http://creativecommons.org/licenses/by/4.0/"><img
alt="Creative Commons License" style="border-width:0"
src="https://i.creativecommons.org/l/by/4.0/88x31.png" />
</a><br />This work is licensed under a <a rel="license"
href="http://creativecommons.org/licenses/by/4.0/">Creativ
e Commons Attribution 4.0 International License</a>.
```

● Normal Icon ○ Compact Icon

Recommended: CC-BY and CC0

Attribution
CC BY

This license lets others distribute, remix, tweak, and build upon your work, even commercially, as long as they credit you for the original creation. This is the most accommodating of licenses offered. Recommended for maximum dissemination and use of licensed materials.

CC0 (datasets) “No rights reserved”

CC0 can be particularly important for the sharing of data and databases, since it otherwise may be unclear whether highly factual data and databases are restricted by copyright or other rights. Databases may contain facts that, in and of themselves, are not protected by copyright law.

CC0 is recommended for data and databases and is used by hundreds of organizations. It is especially recommended for scientific data. Although CC0 doesn't legally require users of the data to cite the source, it does not take away the moral responsibility to give attribution, as is common in scientific research.

http://creativecommons.org/licenses/

```
               Rv1155,aroG,
               icl,Rv1264,thy
2              Rv0223c,lipJ,Rv1
115            25      cyp130,Rv
20             TB31.7,Rv1264,mscL
1              fabG1,
13             mmaA4,bphD,Rv1264,m
18             TB31.7,cyp130,aroG,
5              pth,ethR,clpP,glbN,
14             pknD,lipJ,fabH,Rv1
10             mmaA4,Rv1264,groE
2              mmaA4,Rv1264,thy
               pepD,Rv1264,thy
               pknD,pepD,fa
```

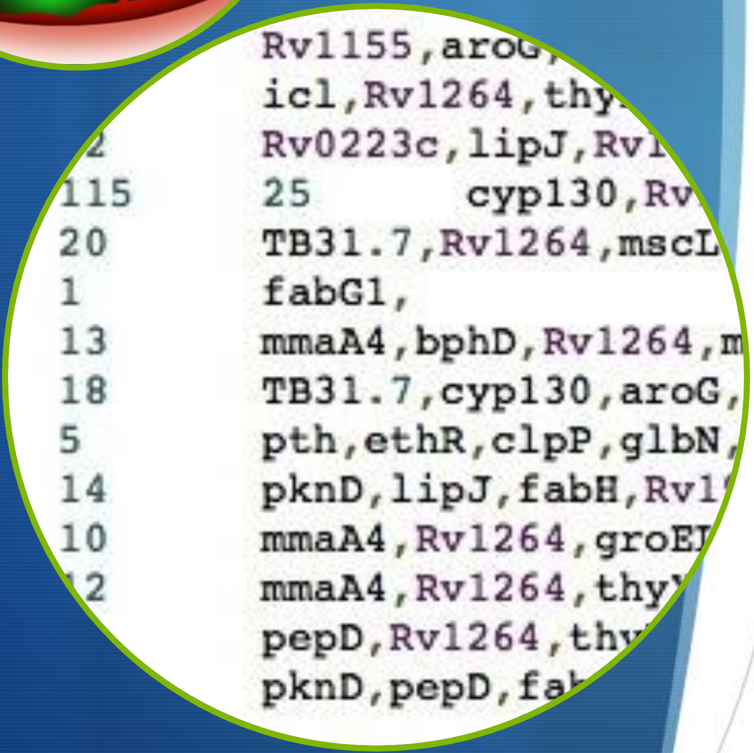
Simplest Approach

1. Create a public entry for your dataset with a persistent unique identifier
   - Go to zenodo.org, create an account
   - Create an entry for your dataset
2. Specify the metadata
   - Including license -- choose from http://www.creativecommons.org/licenses
3. Upload/point to the data

Voilà! Figshare will give you a data citation

1 Data accessibility

Ideal Approach



```
        Rv1155,aroG,
        icl,Rv1264,thy
2       Rv0223c,lipJ,Rv1
115     25      cyp130,Rv
20      TB31.7,Rv1264,mscL
1       fabG1,
13      mmaA4,bphD,Rv1264,m
18      TB31.7,cyp130,aroG,
5       pth,ethR,clpP,glbN,
14      pknD,lipJ,fabH,Rv1
10      mmaA4,Rv1264,groEL
12      mmaA4,Rv1264,thy
        pepD,Rv1264,thy
        pknD,pepD,fa
```

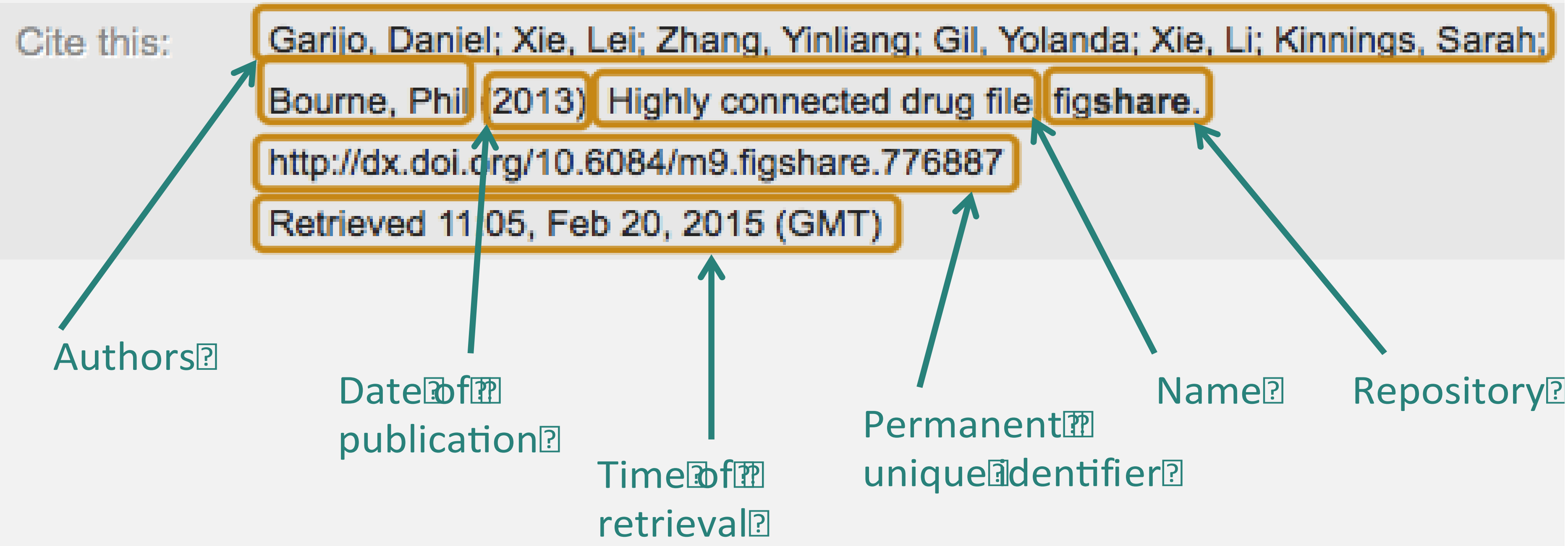
1. Find a repository that your community uses, if there is not one then organize one!
2. Create a public entry for your dataset with a persistent unique identifier
   - Create an entry for your dataset
3. Specify the metadata required by that repository using metadata standards for that community
   - Including license -- choose from http://www.creativecommons.org/licenses
4. Upload/point to the data
5. Get a data citation from the repository

1 Data accessibility

What to Show in a GPF

- ★ Cite each of your datasets like you would cite another paper
- ★ Citation includes publication date, date of retrieval, repository, and persistent identifier
- ★ If there is a data paper, cite it

Data Citation Format



Cite this: Garijo, Daniel; Xie, Lei; Zhang, Yinliang; Gil, Yolanda; Xie, Li; Kinnings, Sarah; Bourne, Phil (2013) Highly connected drug file figshare. http://dx.doi.org/10.6084/m9.figshare.776887 Retrieved 11:05, Feb 20, 2015 (GMT)

Authors

Date of publication

Time of retrieval

Permanent unique identifier

Name

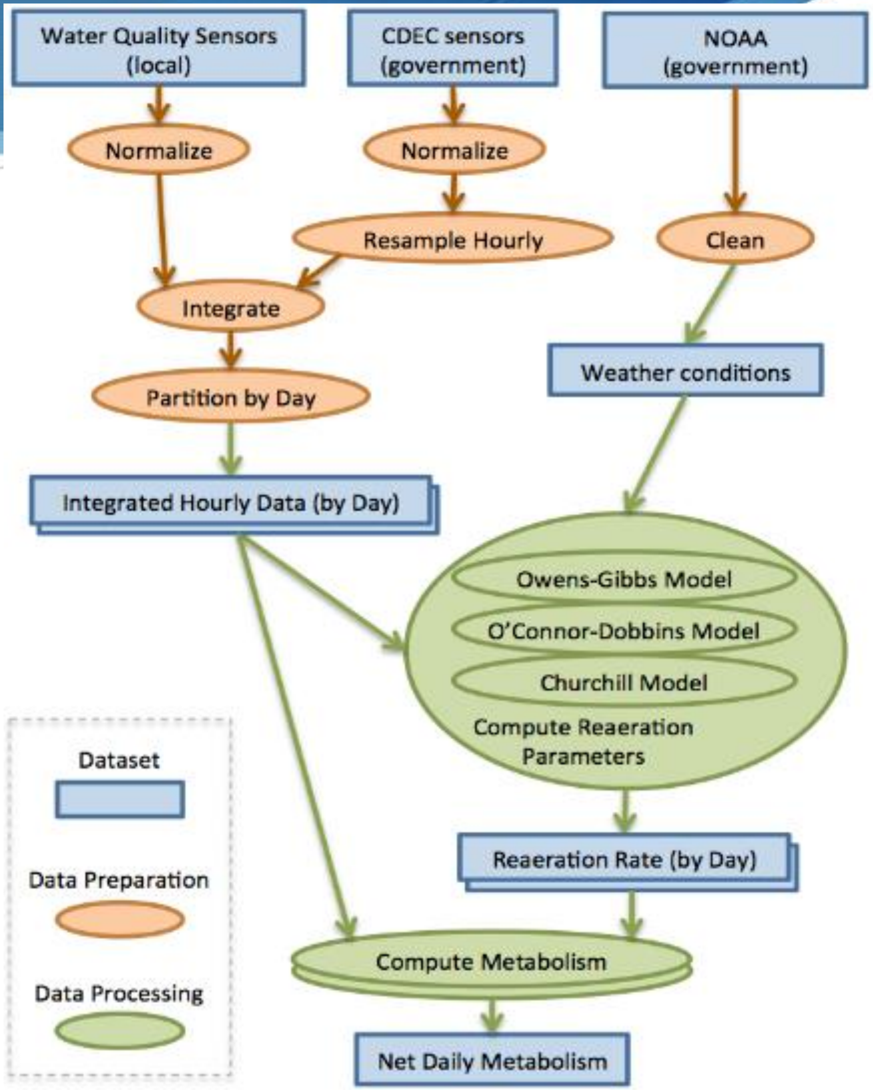
Repository

5 Provenance documentation

6 Methods documentation

How to Sketch a Workflow

1. Compile the command line invocation to all your codes
   - ★ Input data, parameters, configuration files
   - ★ Include data preparation codes
2. Consider how the data flows from code to code
3. Starting with the input data, work your way to the results
4. If any steps were done with manual intervention, indicate that
5. Create subworkflows if it gets large



2 Data documentation

<MMI/> Marine Metadata Interoperability

CF MetaData

ISO 19115

WaterML2.0

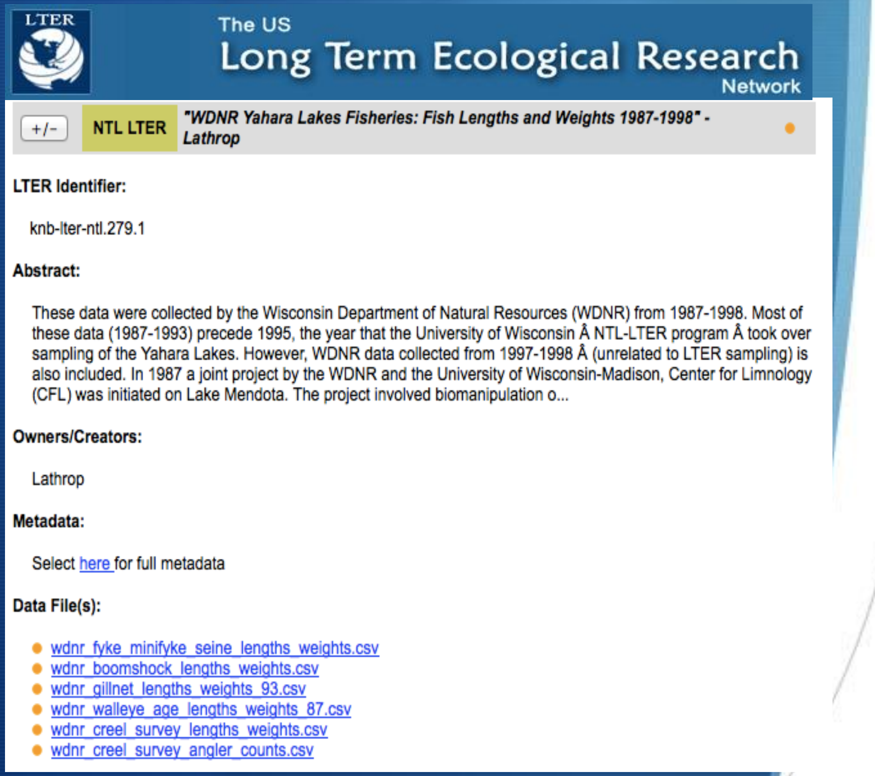
Simplest Approach

- ★ Datasets should have general-purpose metadata specified (creator, date, name, etc.)

Ideal Approach

- ★ Dataset characteristics should be explained in detail
- ★ Domain-specific metadata should be documented
- ★ Availability of related datasets should be documented

What to Show in the Paper



NTL LTER "WDNR Yahara Lakes Fisheries: Fish Lengths and Weights 1987-1998" - *Lathrop*

LTER Identifier:

knb-lter-ntl.279.1

Abstract:

These data were collected by the Wisconsin Department of Natural Resources (WDNR) from 1987-1998. Most of these data (1987-1993) precede 1995, the year that the University of Wisconsin Â NTL-LTER program Â took over sampling of the Yahara Lakes. However, WDNR data collected from 1997-1998 Â (unrelated to LTER sampling) is also included. In 1987 a joint project by the WDNR and the University of Wisconsin-Madison, Center for Limnology (CFL) was initiated on Lake Mendota. The project involved biomanipulation o...

Owners/Creators:

Lathrop

Metadata:

Select here for full metadata

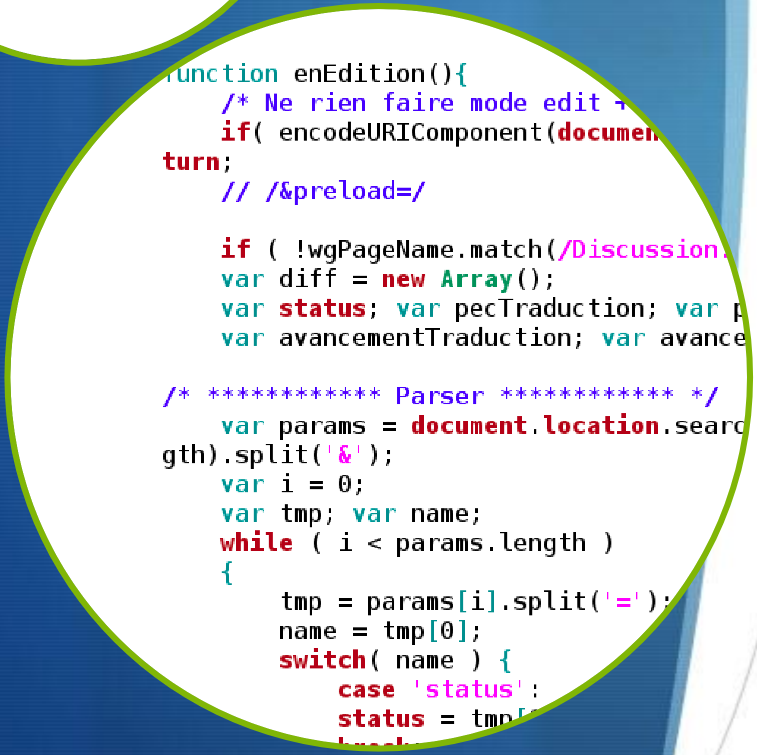
Data File(s):

- wdnr_fyke_minifyke_seine_lengths_weights.csv
- wdnr_boomshock_lengths_weights.csv
- wdnr_gillnet_lengths_weights_93.csv
- wdnr_walleye_age_lengths_weights_87.csv
- wdnr_creel_survey_lengths_weights.csv
- wdnr_creel_survey_angler_counts.csv

★ Mention that the persistent identifier for your data has pointers to its metadata and includes a detailed description of the data

★ Optionally, include the metadata also as supplemental material

★ If there is a data paper, cite it



```
unction enEdition(){
    /* Ne rien faire mode edit +
    if( encodeURIComponent(documen
turn;
    // /&preload=/

    if ( !wgPageName.match(/Discussion
    var diff = new Array();
    var status; var pecTraduction; var p
    var avancementTraduction; var avance

/* ************ Parser ************ */
    var params = document.location.searc
gth).split('&');
    var i = 0;
    var tmp; var name;
    while ( i < params.length )
    {
        tmp = params[i].split('=');
        name = tmp[0];
        switch( name ) {
            case 'status':
            status = tmp[
```

https://commons.wikimedia.org/wiki/File:Source_code_in_Javascript.png

Simplest Approach

1. Create a public entry for your software with a persistent unique identifier
   - Post on your web site and use a PURL, upload to figshare as you would data and get a DOI
2. Specify the metadata
   - Including license -- choose from http://opensource.org/licenses, preferably Apache v2.0
3. Specify desired citation

```
function enEdition(){
    /* Ne rien faire mode edit +
    if( encodeURIComponent(documen
turn;
    // /&preload=/

    if ( !wgPageName.match(/Discussion
    var diff = new Array();
    var status; var pecTraduction; var p
    var avancementTraduction; var avance

/* ************ Parser ************ */
    var params = document.location.searc
gth).split('&');
    var i = 0;
    var tmp; var name;
    while ( i < params.length )
    {
        tmp = params[i].split('=');
        name = tmp[0];
        switch( name ) {
            case 'status':
            status = tmp[
```

https://commons.wikimedia.org/wiki/File:Source_code_in_Javascript.png

Ideal Approach

1. Learn to use a code repository that allows version tracking and collaborative software development
    - GitHub, BitBucket, etc.
2. Create a public entry for your software with a persistent unique identifier
3. Specify the metadata
    - Including license -- choose from http://opensource.org/licenses, preferably Apache v2.0
4. Specify desired citation

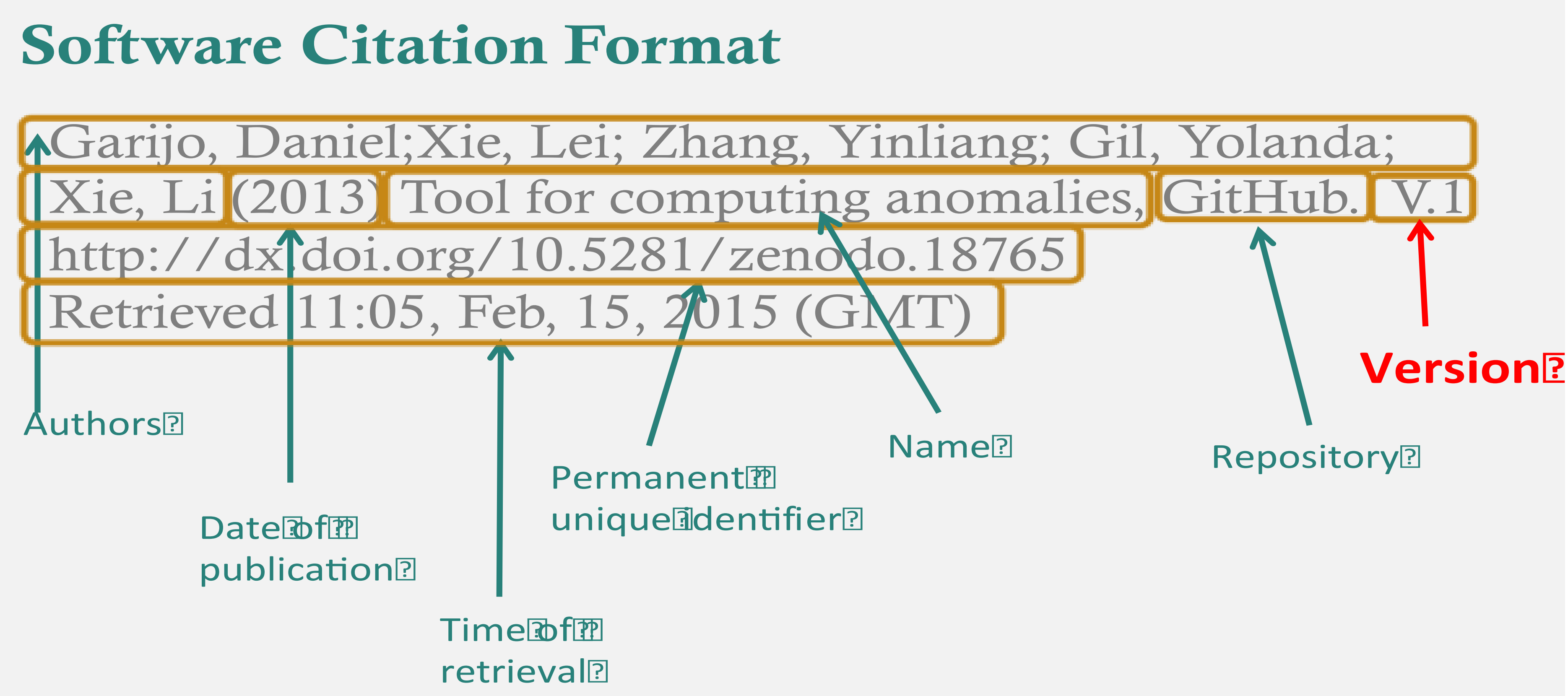
Choosing an Open Source License

- Copyright: automatically applied to software when it is created to grant *the creator* exclusive rights as an intellectual property
- **Open source license**: reduce constraints and enable software developers to make their source code available to public
  1. "Copyleft" license (ex: GNU General Public License (GPL))
  2. "Permissive" license (ex: Apache 2 or MIT licenses)
- **Open Source Initiative**
  - Choose a license from: http://opensource.org/licenses
  - Recommend that you choose a permissive license
    - Apache v2

3 Software accessibility

What to Show in a GPF

- Cite each piece of software that you use (preparation, analysis, visualization) like you would cite another paper
  - Citation similar to data but includes software version
- If there is a software paper, cite it

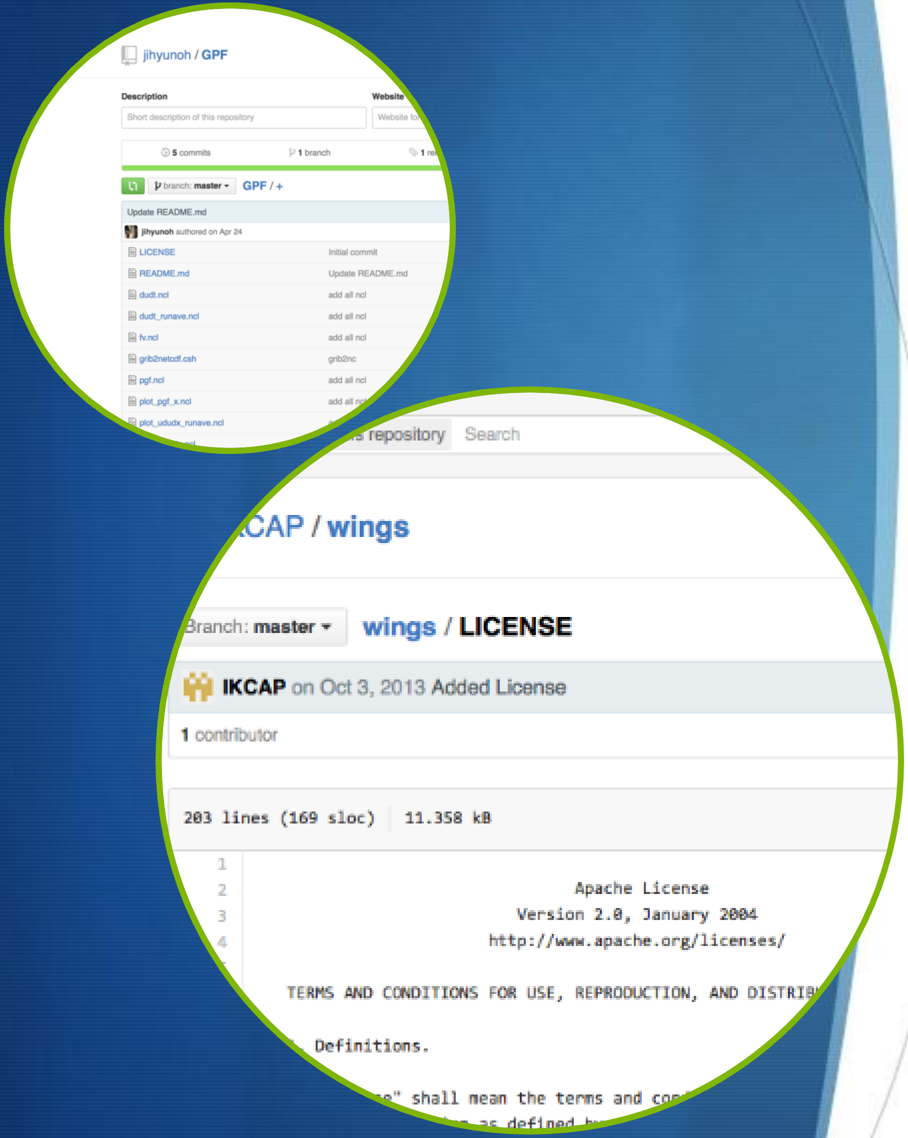


Software Citation Format

Garijo, Daniel;Xie, Lei; Zhang, Yinliang; Gil, Yolanda; Xie, Li (2013) Tool for computing anomalies, GitHub. V.1 http://dx.doi.org/10.5281/zenodo.18765 Retrieved 11:05, Feb, 15, 2015 (GMT)

Simplest Approach

1. Describe as much metadata as you can in your software site
   1. Document the basic metadata
   2. If you use a code repository, there is some basic structure you can follow



jihyunoh / GPF

Description

Short description of this repository

Website

Website fo

5 commits | 1 branch | 1 re

branch: master | GPF / +

Update README.md

jihyunoh authored on Apr 24

LICENSE	Initial commit
README.md	Update README.md
dudt.ncl	add all ncl
dudt_runave.ncl	add all ncl
fv.ncl	add all ncl
grib2netcdf.csh	grib2nc
pgf.ncl	add all ncl
plot_pgf_x.ncl	add all nc
plot_ududx_runave.ncl	

s repository | Search

CAP / wings

Branch: master | wings / LICENSE

IKCAP on Oct 3, 2013 Added License

1 contributor

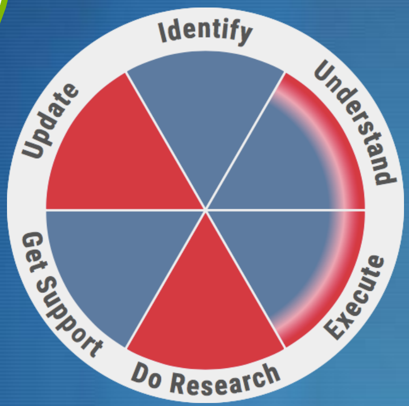
203 lines (169 sloc) | 11.358 kB

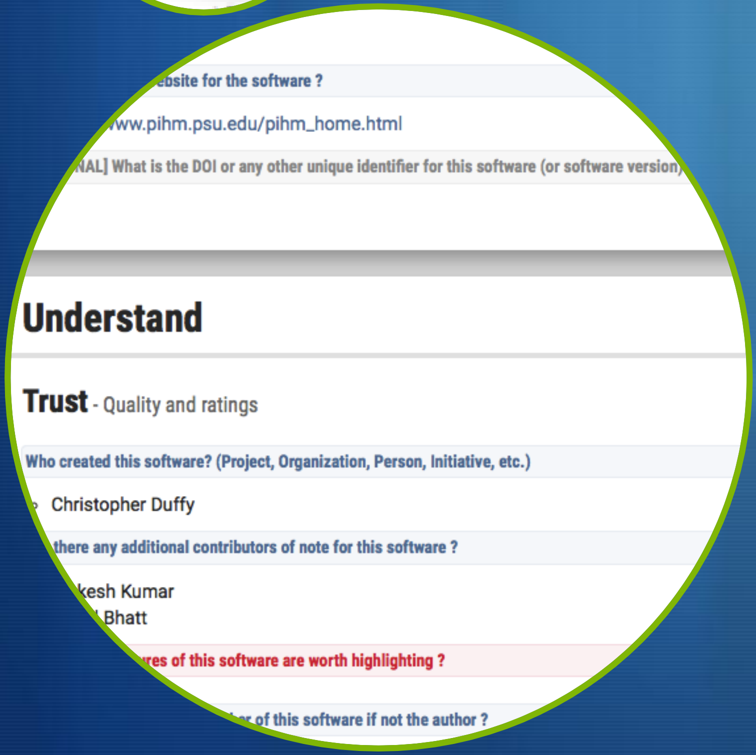
```
                              Apache License
                        Version 2.0, January 2004
                     http://www.apache.org/licenses/

TERMS AND CONDITIONS FOR USE, REPRODUCTION, AND DISTRIB

. Definitions.

 e" shall mean the terms and con
 ... as defined by
```

4 Software documentation

Ideal Approach

1. Use software registry
   - http://www.ontosoft.org/portal, csdms.colorado.edu, etc.
   - Guides through questions to provide metadata
2. Save the metadata as HTML, XML,…
3. Post the metadata on your code site




ebsite for the software ?

www.pihm.psu.edu/pihm_home.html

NAL] What is the DOI or any other unique identifier for this software (or software version)

Understand

Trust - Quality and ratings

Who created this software? (Project, Organization, Person, Initiative, etc.)

Christopher Duffy

there any additional contributors of note for this software ?

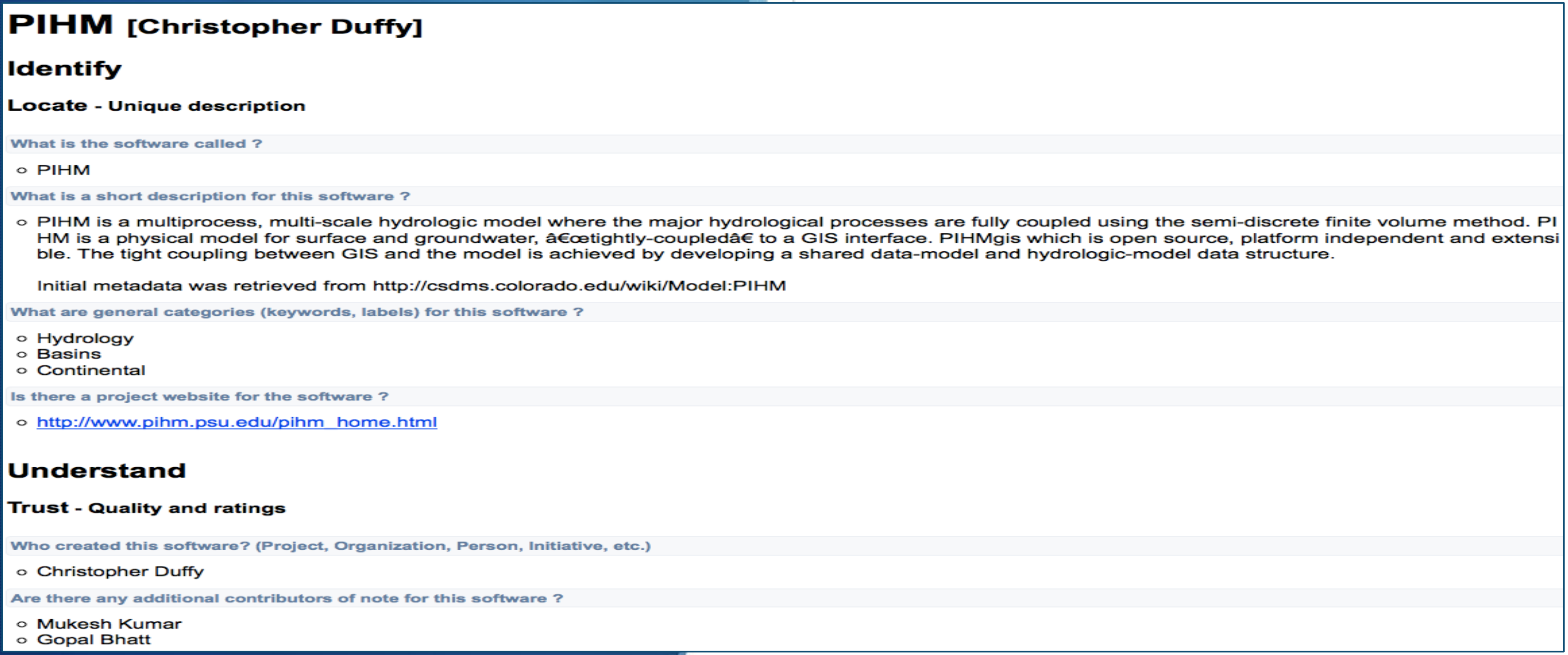
kesh Kumar
Bhatt

res of this software are worth highlighting ?

r of this software if not the author ?

What to Show in the Paper

- ★ Mention that the persistent identifier location for your software points to its metadata
- ★ Optionally, include the software metadata as supplemental material
- ★ If there is a software paper, cite it



PIHM [Christopher Duffy]

Identify

Locate - Unique description

What is the software called ?

- PIHM

What is a short description for this software ?

- PIHM is a multiprocess, multi-scale hydrologic model where the major hydrological processes are fully coupled using the semi-discrete finite volume method. PIHM is a physical model for surface and groundwater, â€œtightly-coupledâ€ to a GIS interface. PIHMgis which is open source, platform independent and extensible. The tight coupling between GIS and the model is achieved by developing a shared data-model and hydrologic-model data structure.

  Initial metadata was retrieved from http://csdms.colorado.edu/wiki/Model:PIHM

What are general categories (keywords, labels) for this software ?

- Hydrology
- Basins
- Continental

Is there a project website for the software ?

- http://www.pihm.psu.edu/pihm_home.html

Understand

Trust - Quality and ratings

Who created this software? (Project, Organization, Person, Initiative, etc.)

- Christopher Duffy

Are there any additional contributors of note for this software ?

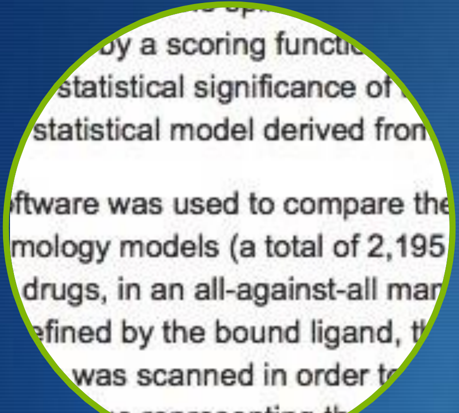
- Mukesh Kumar
- Gopal Bhatt

5 Provenance documentation

6 Methods documentation

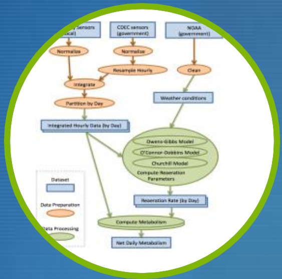
Simplest Approach

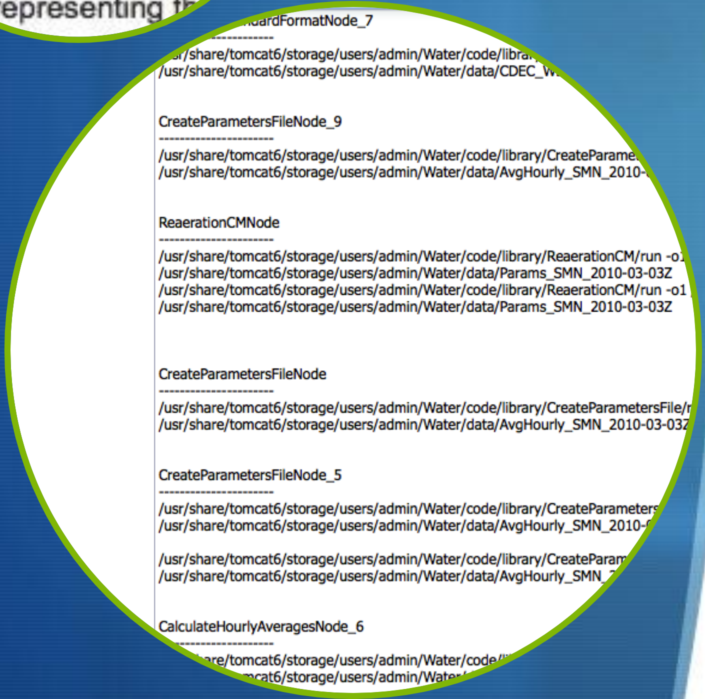
1. Describe the workflow in text
   - Data + software + workflow
   - Specify unique identifiers for data and software, versions, credit all sources
2. Develop a workflow sketch
   - Capture high-level dataflow across components
3. For provenance, include a summary or an execution trace



by a scoring functi
statistical significance of
statistical model derived fron

ftware was used to compare the
mology models (a total of 2,195
drugs, in an all-against-all mar
efined by the bound ligand, t
was scanned in order t
e representing th





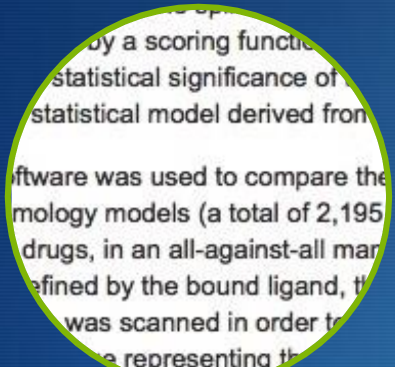
```
ardFormatNode_7
----------------------
r/share/tomcat6/storage/users/admin/Water/code/librar
/usr/share/tomcat6/storage/users/admin/Water/data/CDEC_W

CreateParametersFileNode_9
----------------------
/usr/share/tomcat6/storage/users/admin/Water/code/library/CreateParame
/usr/share/tomcat6/storage/users/admin/Water/data/AvgHourly_SMN_2010-

ReaerationCMNode
----------------------
/usr/share/tomcat6/storage/users/admin/Water/code/library/ReaerationCM/run -o1
/usr/share/tomcat6/storage/users/admin/Water/data/Params_SMN_2010-03-03Z
/usr/share/tomcat6/storage/users/admin/Water/code/library/ReaerationCM/run -o1
/usr/share/tomcat6/storage/users/admin/Water/data/Params_SMN_2010-03-03Z


CreateParametersFileNode
----------------------
/usr/share/tomcat6/storage/users/admin/Water/code/library/CreateParametersFile/r
/usr/share/tomcat6/storage/users/admin/Water/data/AvgHourly_SMN_2010-03-03Z

CreateParametersFileNode_5
----------------------
/usr/share/tomcat6/storage/users/admin/Water/code/library/CreateParameters
/usr/share/tomcat6/storage/users/admin/Water/data/AvgHourly_SMN_2010-

/usr/share/tomcat6/storage/users/admin/Water/code/library/CreateParam
/usr/share/tomcat6/storage/users/admin/Water/data/AvgHourly_SMN_2

CalculateHourlyAveragesNode_6
----------------------
are/tomcat6/storage/users/admin/Water/code/l
cat6/storage/users/admin/Water/
```

5 Provenance documentation

6 Methods documentation

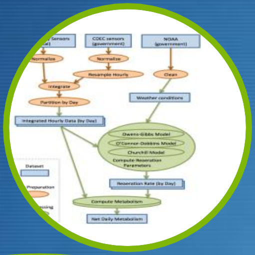
Ideal Approach

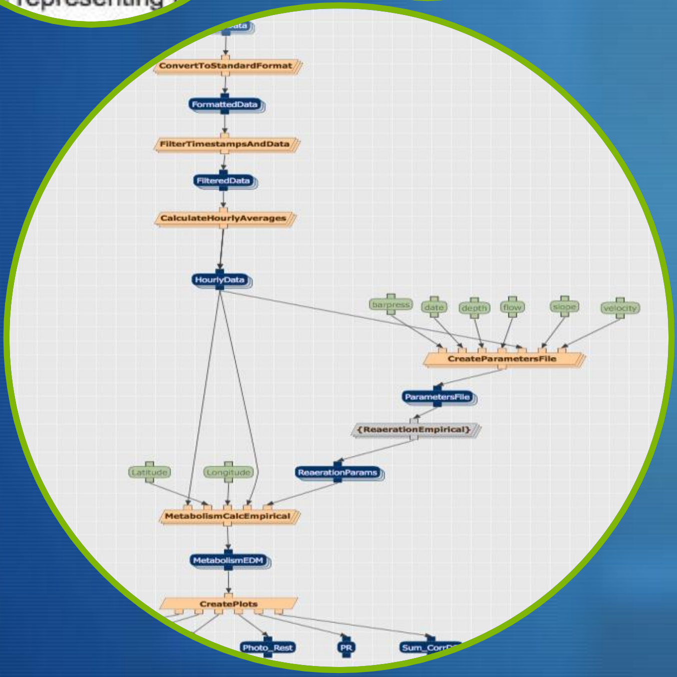
1. Describe the workflow in text
   - Data + software + workflow
   - Specify unique identifiers for data and software, versions, credit all sources
2. Develop a workflow sketch
   - Capture high-level dataflow across components
3. Specify the formal workflow using a workflow system, electronic notebook, etc.
   - Command lines + parameter values
   - Dataflow across components
4. Include the provenance record
   - If generating it automatically, preferably using a standard (e.g., PROV)
5. Publish the workflow and provenance record in a publicly accessible repository (eg figshare, myExperiment, etc)
6. Get a unique persistent identifier for the workflow, the provenance, or both



by a scoring functi
statistical significance of
statistical model derived fron

ftware was used to compare the
mology models (a total of 2,195
drugs, in an all-against-all ma
efined by the bound ligand, t
was scanned in order t
e representing th



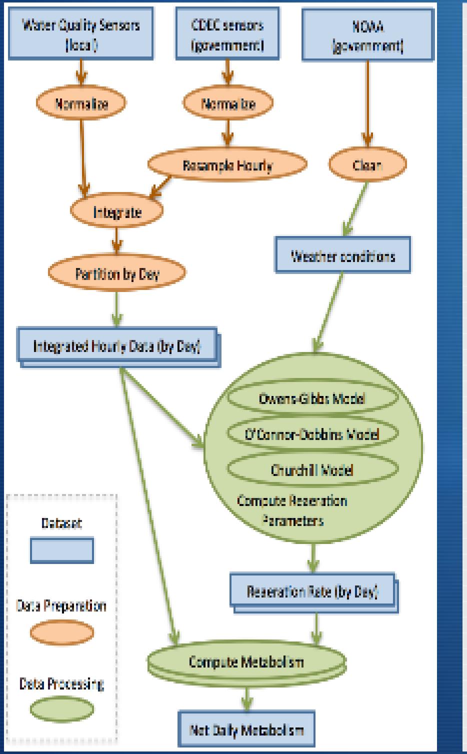


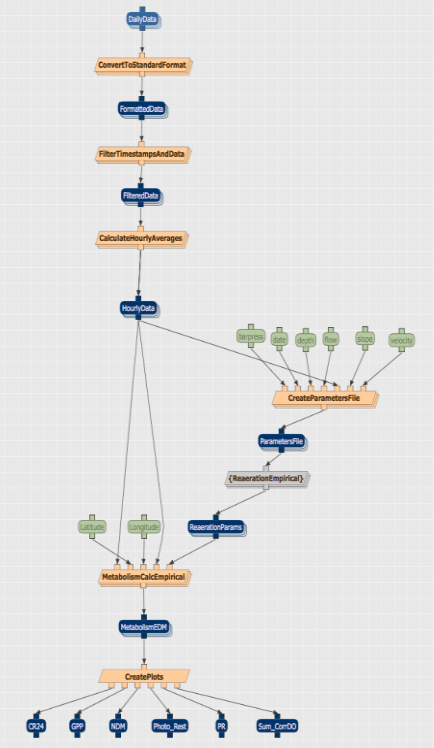
5 Provenance documentation

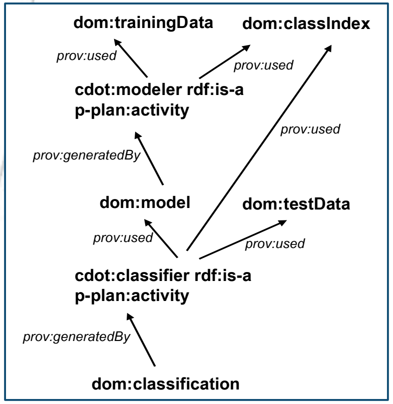
6 Methods documentation

What to Show in the Paper

- ★ Describe workflow in text and provide a workflow sketch
  - ★ Optionally, provide the formal workflow or lab notebook, use a persistent identifier, and cite it
- ★ Include a summary of the execution traces as supplementary material, or use a persistent identifier and cite it
  - ★ Optionally, include instead the provenance records using a standard like W3C PROV






dom:trainingData

dom:classIndex

prov:used

prov:used

cdot:modeler rdf:is-a p-plan:activity

prov:used

prov:generatedBy

dom:model

dom:testData

prov:used

prov:used

cdot:classifier rdf:is-a p-plan:activity

prov:generatedBy

dom:classification

```
# Entities
ex:testData1 a prov:Entity .
ex:model1 a prov:Entity .
ex:classification1 a prov:Entity .
# Activities
ex:Classifier1 a prov:Activity .
# Usage and Generation
relations between entities
and activities
ex:Classifier1
    prov:used ex:testData1 ;
    prov:used ex:model1 .
ex:classification1
    prov:wasGeneratedBy
        ex:Classifier1 .
```

7 Authors identification

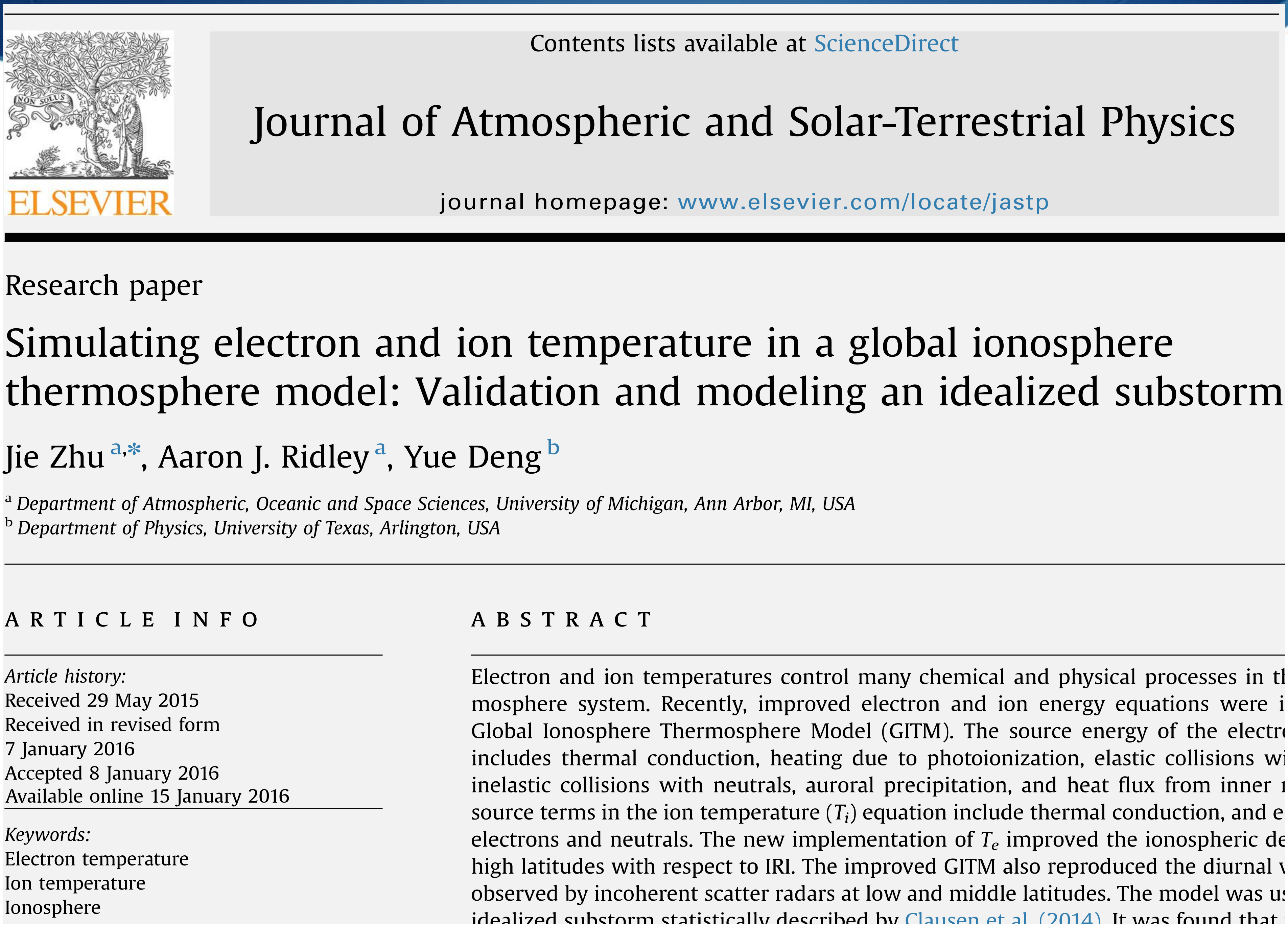
What to Show in the Paper

https://pixabay.com/en/people-character-faces-real-3058361/

- ★ Authors have a persistent unique identifier
  - ★ Use www.orcid.org

ORCID

EXTRA SLIDES: Workflows from [Zhu, Ridley, and Deng 2016]



Contents lists available at ScienceDirect

Journal of Atmospheric and Solar-Terrestrial Physics

journal homepage: www.elsevier.com/locate/jastp

Research paper

Simulating electron and ion temperature in a global ionosphere thermosphere model: Validation and modeling an idealized substorm

Jie Zhu[a,*], Aaron J. Ridley[a], Yue Deng[b]

[a] *Department of Atmospheric, Oceanic and Space Sciences, University of Michigan, Ann Arbor, MI, USA*
[b] *Department of Physics, University of Texas, Arlington, USA*

ARTICLE INFO

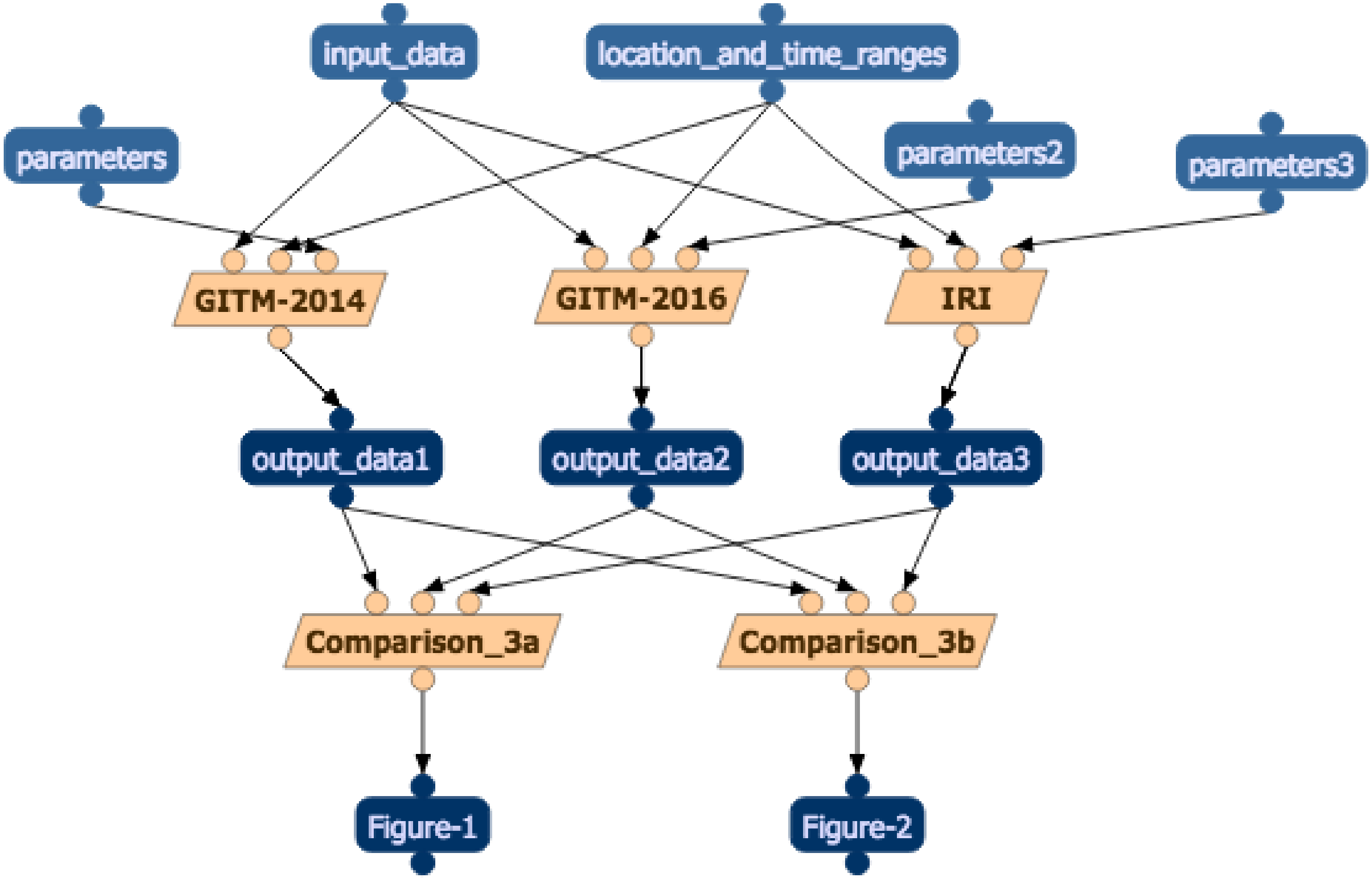
Article history:
Received 29 May 2015
Received in revised form
7 January 2016
Accepted 8 January 2016
Available online 15 January 2016

Keywords:
Electron temperature
Ion temperature
Ionosphere

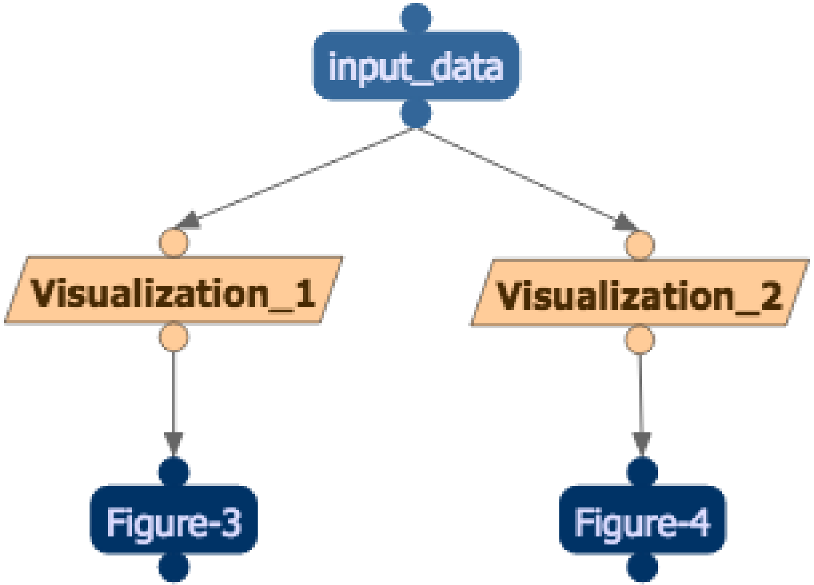
ABSTRACT

Electron and ion temperatures control many chemical and physical processes in tl
mosphere system. Recently, improved electron and ion energy equations were i
Global Ionosphere Thermosphere Model (GITM). The source energy of the electr
includes thermal conduction, heating due to photoionization, elastic collisions wi
inelastic collisions with neutrals, auroral precipitation, and heat flux from inner r
source terms in the ion temperature (T_i) equation include thermal conduction, and e
electrons and neutrals. The new implementation of T_e improved the ionospheric de
high latitudes with respect to IRI. The improved GITM also reproduced the diurnal v
observed by incoherent scatter radars at low and middle latitudes. The model was us
idealized substorm statistically described by Clausen et al. (2014). It was found that

Figures 1 & 2


input_data
location_and_time_ranges
parameters
parameters2
parameters3
GITM-2014
GITM-2016
IRI
output_data1
output_data2
output_data3
Comparison_3a
Comparison_3b
Figure-1
Figure-2

Figures 3 & 4


input_data
Visualization_1
Visualization_2
Figure-3
Figure-4

Figures 5 & 6 & 9

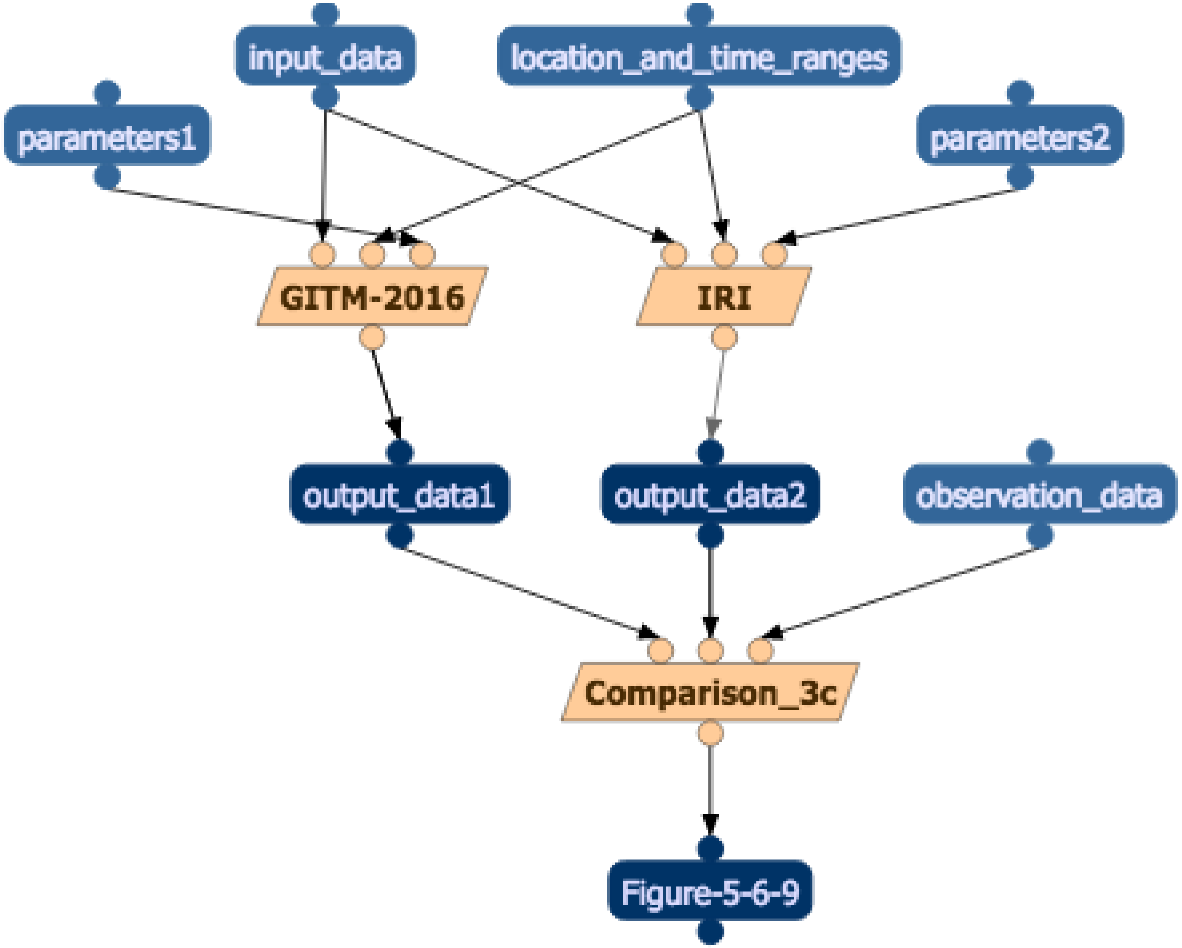

input_data
location_and_time_ranges
parameters1
parameters2
GITM-2016
IRI
output_data1
output_data2
observation_data
Comparison_3c
Figure-5-6-9

Figure 7

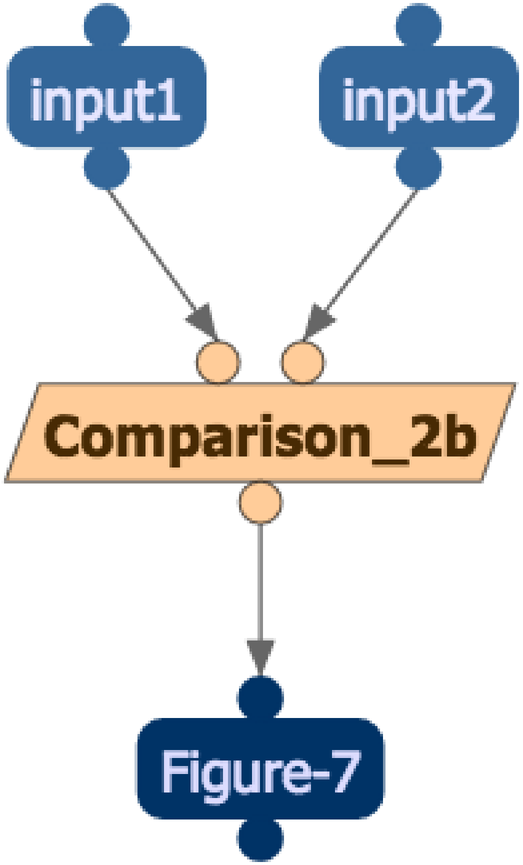
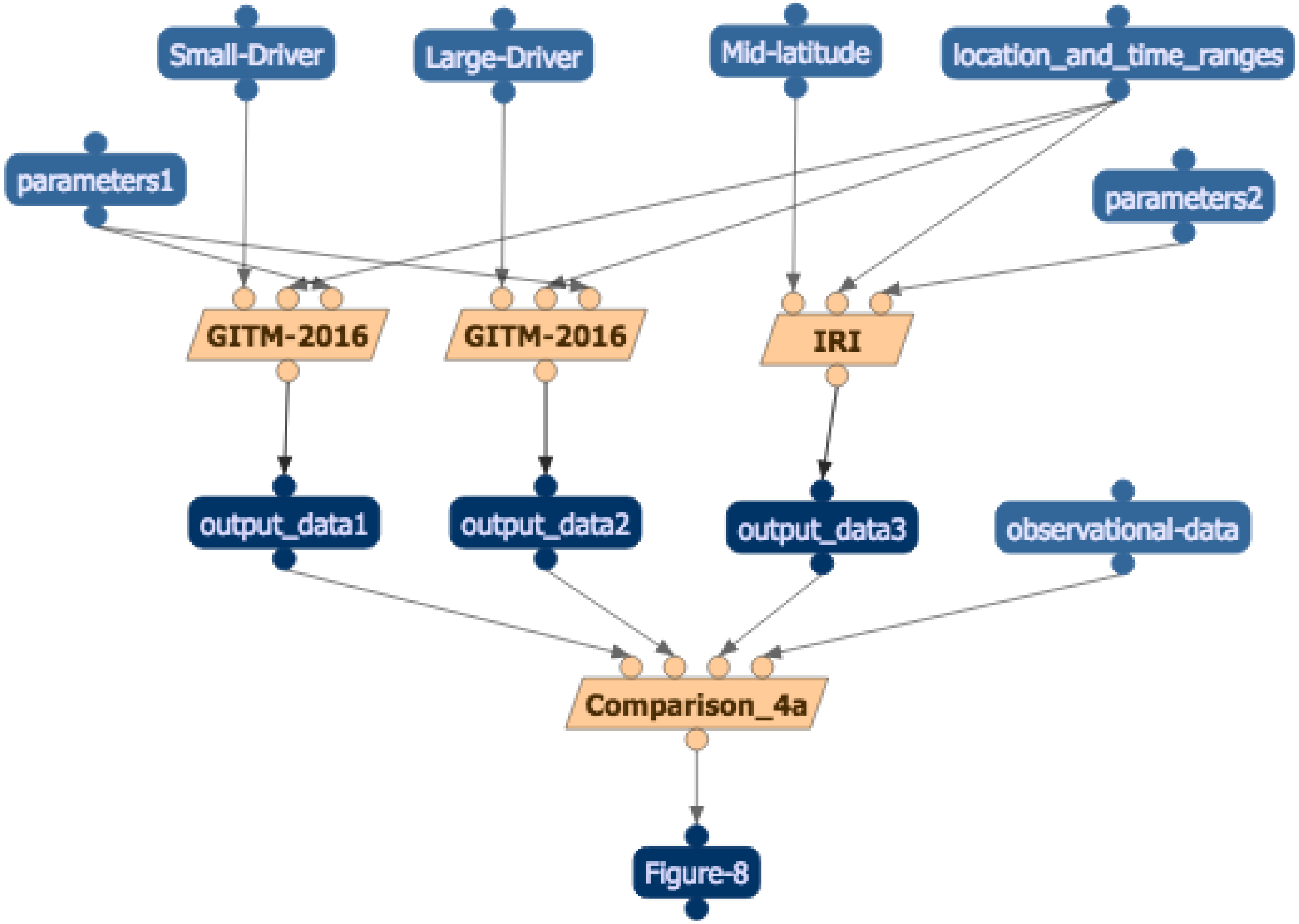
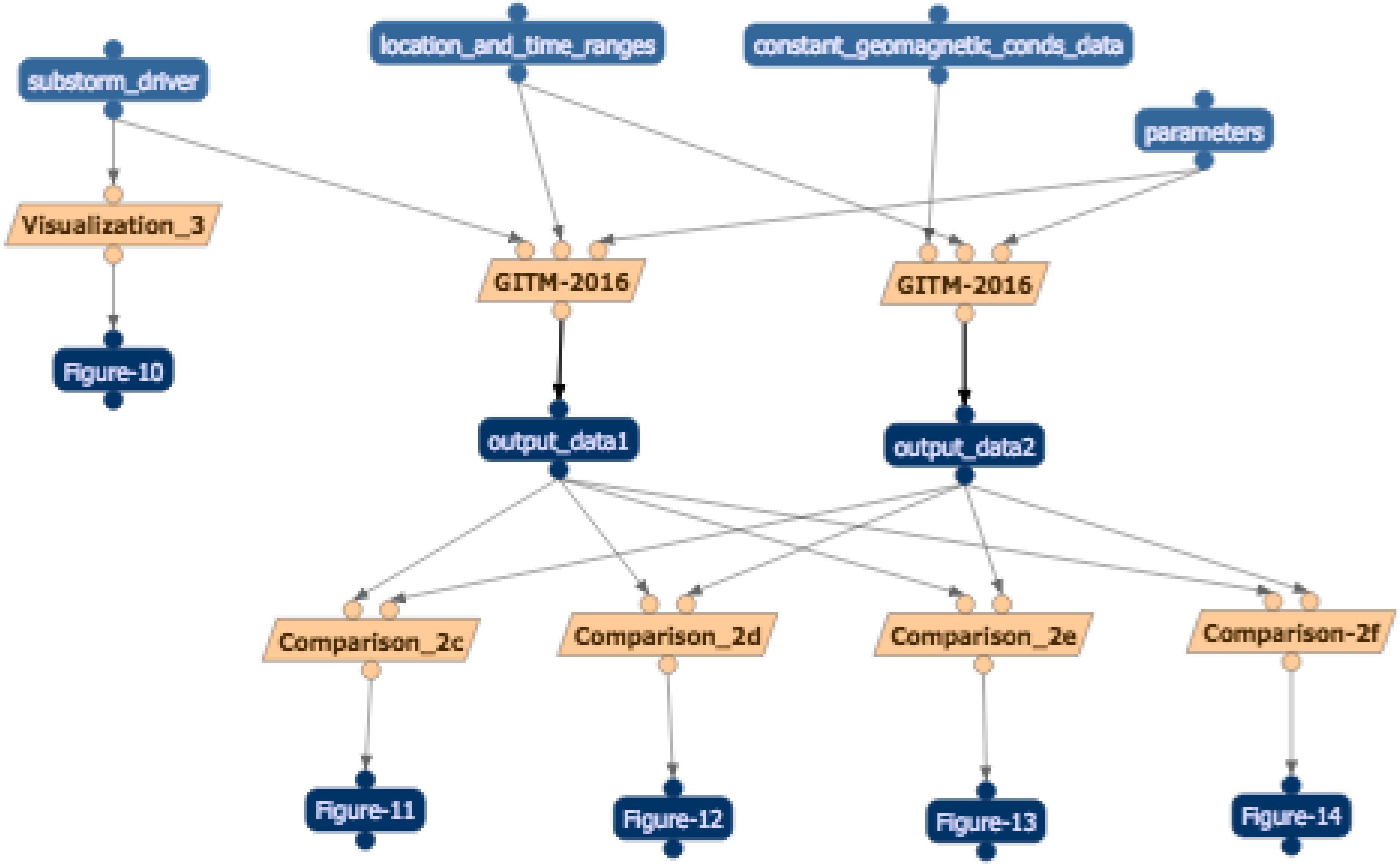


Figure 8


Small-Driver
Large-Driver
Mid-latitude
location_and_time_ranges
parameters1
parameters2
GITM-2016
GITM-2016
IRI
output_data1
output_data2
output_data3
observational-data
Comparison_4a
Figure-8

Figures 11 & 12 & 13 & 14


location_and_time_ranges
constant_geomagnetic_conds_data
substorm_driver
parameters
Visualization_3
GITM-2016
GITM-2016
Figure-10
output_data1
output_data2
Comparison_2c
Comparison_2d
Comparison_2e
Comparison-2f
Figure-11
Figure-12
Figure-13
Figure-14