

11 February 2021 | 100

# Science

change  
use  
information  
climate  
science  
model  
need  
sharing  
brain  
challenges  
analysis  
visualization  
new  
many  
researchers  
access  
example  
knowledge  
data

nature

THE BITER BIT  
Wind turbines for insects

TROPICAL CYCLONES  
The strong get stronger

BLACK HOLE PHYSICS  
A new window on the Galactic Centre

**BIG DATA**

24 SEPTEMBER 2012  
£10.00 / \$18.00

**SCIENCE IN THE PETABYTE ERA**

**The Economist**

Obama: the winner  
Misgoverning Argentina  
The economic shift from West to East  
Somewhat modified crops blossom  
The right to eat cats and dogs

**The data deluge**  
AND HOW TO HANDLE IT: A 14-PAGE SPECIAL REPORT

A man in a dark suit and white shirt is standing, holding a large green umbrella over his head. He is also holding a watering can and watering a small, stylized plant that has a single orange flower. The background is white with faint, vertical lines of text, suggesting a data or information theme.

# My Background (short extract)

- CAU Kiel, Department of Computer Science, since 2008
  - Head, Software Engineering Group
  - Previously: University of Oldenburg (similar position 2000-2008), Tilburg (NL), Dortmund, Essen, Braunschweig.
- Competence cluster  
Software Systems Engineering (KoSSE)
- Innovationsstiftung  
Schleswig-Holstein
- Excellence cluster “Future Ocean”
  - Principal Investigator, Research Area R10 “Ocean Observations”  
future ocean  
KIEL MARINE SCIENCES
- Currently: Dean, Faculty of Engineering

# A brief look at Software Engineering Terminology

- Software:
  - “Computer programs, procedures, rules, and possibly associated documentation and data pertaining to the operation of a computer system.”
  - IEEE Standard Glossary of Software Engineering Terminology, IEEE Standard 610.12-1990
- Concerning the term “model”:
  - Software as model for the real world
  - Model of the Software
- To distinguish:
  - Software as Code
  - Software as a Service

# Agenda

- What's the problem / challenge that I'm talking about?
  - And what I'm not talking about today ...
- What's the current state?
  - And what you could already do ...
- What to expect in the future?
  - And what you could do ...

# Recommendations of the Commission on Professional Self Regulation in Science

---

Proposals for Safeguarding Good Scientific Practice  
January, 1998



## Recommendation 7 (of 16):

- Primary data as the basis for publications shall be securely stored for ten years in a durable form in the institution of their origin.
- Experiments and numerical calculations can only be repeated if all important steps are **reproducible**.  
For this purpose, they must be recorded.

(Source: [http://www.dfg.de/en/research\\_funding/legal\\_conditions/good\\_scientific\\_practice/index.html](http://www.dfg.de/en/research_funding/legal_conditions/good_scientific_practice/index.html))

**“If I have seen further it is by standing on the shoulders of giants.”**

Isaac Newton, 1676

# Not only in Germany ...

EXECUTIVE DE  
OFFICE OF

## OPEN ACCESS RESEARCH DATA WITHOUT BARRIERS

Research Data Alliance Launch and First Plenary  
March 18-20, 2013, Gothenburg, Sweden

"Sharing and cooperation are essential to science – no wonder scientists have long sought out tools to help them do this better. Remember it was scientists at CERN who invented the World Wide Web. That was a great gift of science to society: now we can ensure that it helps the scientists back."

Neelie Kroes, Vice President of the European Commission responsible for the Digital Agenda<sup>1</sup>

### Policy Principles

The Administration is committed to ensuring that, to the greatest extent and with the fewest constraints possible and consistent with law and the objectives set out below, the direct results of federally funded scientific research are made available to and useful for the public, industry, and the scientific community. Such results include peer-reviewed publications and digital data.

<https://www.eff.org/sites/default/files/ostp-public-access-memo.pdf>

# Scientific misconduct and other challenges

- Prof. Dullo presented several examples of scientific misconduct, such as the case Jan Hendrik Schön.
  - Thus, I can skip that part in my presentation.
- However, there are also other challenges to obey the rules of good scientific practice,
  - that are not scientific misconduct.
- Let's take a look at an example from the work of one of my Ph.D. students...

# A Challenge for Arne's PhD research



## Marine Biology Research

Publication details, including instructions for authors and subscription information:

<http://www.tandfonline.com/loi/smar20>

## Estimating the horizontal and temporal overlap of pelagic fish distribution in the Norwegian Sea using individual-based modelling

Kjell Rong Utne<sup>a</sup> & Geir Huse<sup>a</sup>

<sup>a</sup> Institute of Marine Research, Bergen, Norway

Version of record first published: 25 Apr 2012.

<http://dx.doi.org/10.1080/17451000.2011.639781>

- Utne & Huse provide an abstract (in part mathematical) description of their individual-based model, but:
  - We cannot reconstruct the implementation from the provided information
  - Sources for calibration data are named (some are unpublished) but again we cannot reconstruct the specific input data and parameters used.
- Without releasing the source code **and** the input/configuration data of the model, **reproducibility** of the results is hard or even impossible.

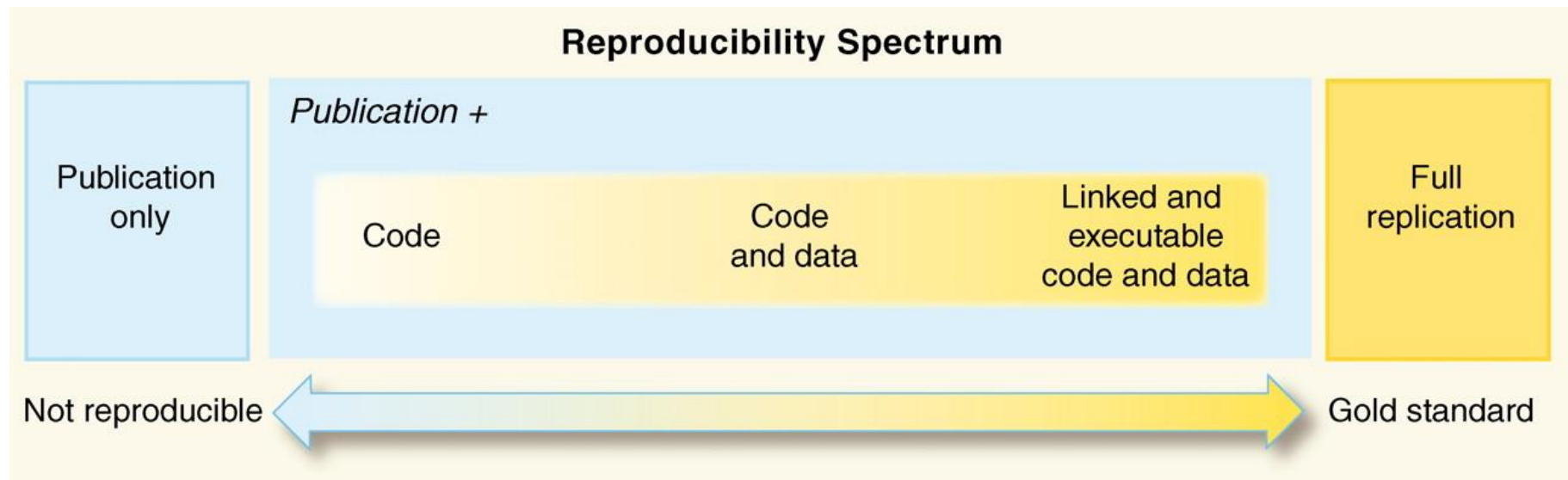


PERSPECTIVE

## Reproducible Research in Computational Science

Roger D. Peng

“Replication is the ultimate standard by which scientific claims are judged.”





## CARING FOR YOUR DATA

By Konrad Hinsen

<http://dx.doi.org/10.1109/MCSE.2012.108>

“Data is at the heart of science.

- A scientist is **expected** to be able to back up all published conclusions with data.
- **Data management** should thus be a priority in science.
- Scientists can’t afford to lose data, be uncertain of what it means, or not know where it came from.
- In the **experimental sciences**, there’s a long tradition of writing down all experimental setups, parameters, and results meticulously in a lab notebook.
- Unfortunately, **computational science** is much less rigorous about data handling, although there are clear signs of improvement.”

# So, what's the problem / challenge that I'm talking about?



- For good scientific practice, it is important that research results may be
  - properly checked by reviewers and
  - possibly repeated and extended by other researchers.
- This is of particular interest for “digital science” i.e. for in silico experiments
- How can Software Systems and Services Contribute?

# What I'm not talking about?

Software and services for detecting plagiarism, such as

[http:// www.plagiarismfinder.de/](http://www.plagiarismfinder.de/)



<http://vroniplag.wikia.com/>

**VroniPlag**

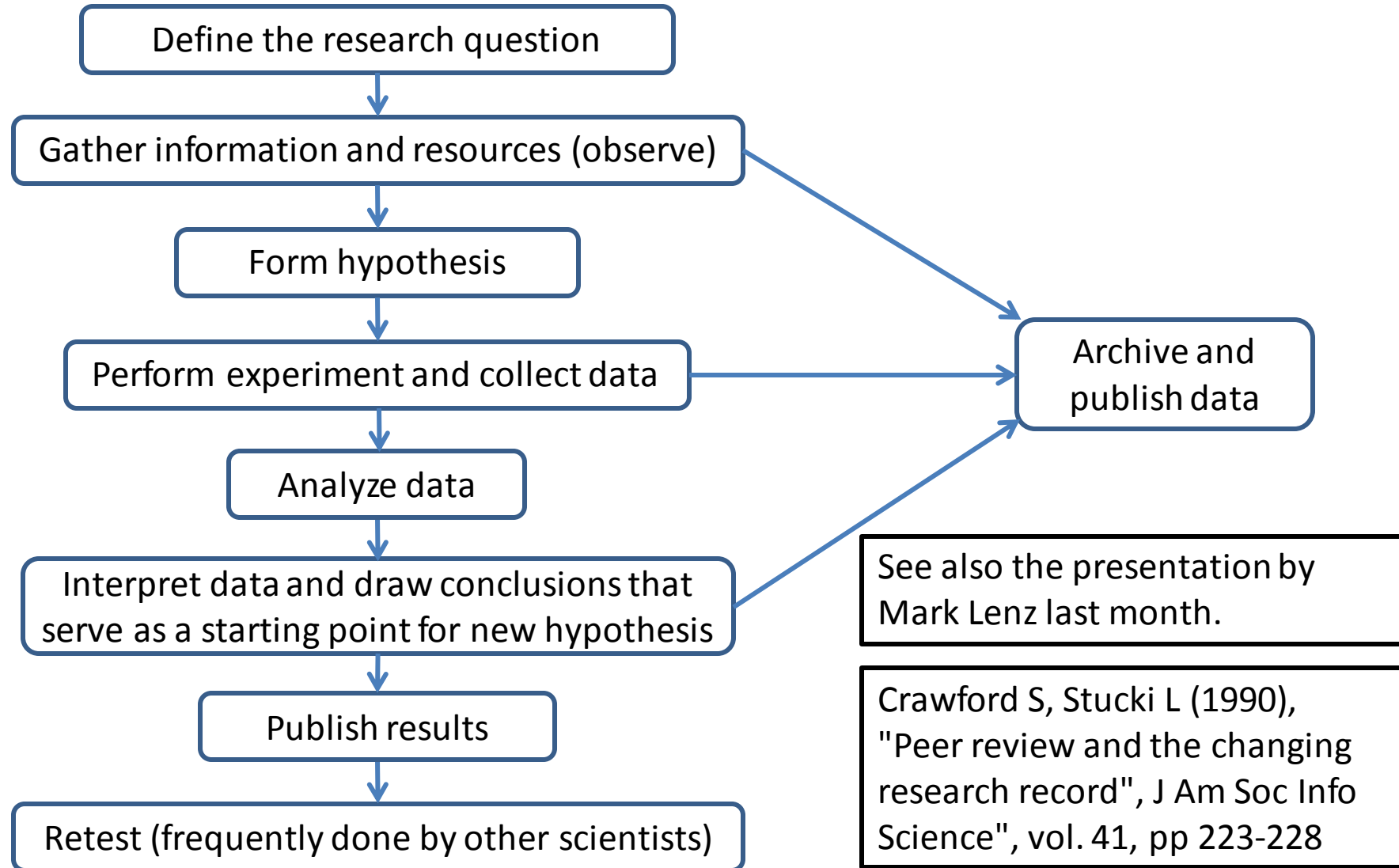
# What I'm not talking about?

- Establishing Software Engineering best practices in computational science, such as
  - Version and configuration management
  - Quality management
  - Software architecture design and modeling
  - Domain-specific programming languages
  - Parallel and distributed programming
- To learn about such topics, you may attend my regular lectures (BSc, MSc) in Computer Science

# Agenda

- What's the problem / challenge that I'm talking about?
  - And what I'm not talking about today ...
- **What's the current state?**
  - **And what you could already do ...**
- What to expect in the future?
  - And what you could do ...

# Research Workflows



# Data Repositories (Services): Examples



## World Data Center for Marine Environmental Sciences

Biogeochemistry, Circulation, and Life of Present and Past Oceans



<http://www.wdc-mare.org/>



**MODEL & DATA**



<http://www.mad.zmaw.de/wdc-for-climate/>



**COMPUTATIONAL  
MATERIALS  
REPOSITORY**



<https://cmr.fysik.dtu.dk/>



# Kiel Data Management Infrastructure – Ocean Science Information System

[SIGN IN](#) [WELCOME](#) [OSIS-KIEL / EXPEDITION-METADATA](#) [ABOUT US](#) [DATA MANAGEMENT \(INTRANET\)](#) [DATA PORTALS](#)

[SIGN IN](#)

[Overview](#) [Legs/Phases](#) [Numerical Models](#) [Other...](#)

Context:

Welcome to Kiel's **Ocean Science Information System** for Cruises/Legs, Expeditions, Model-Experiments...

You can select one of the actions below to start your work in the portal. Notice, that most of the metadata is public and available to everyone. Access to uploaded files, data, etc. may be restricted but a link or other information whom to contact for gaining access should always be provided. If you feel that access restrictions should not apply and something seems odd please contact the data management team directly by phone (+49(0)431 600 4025) or mail to [datamanagement@geomar.de](mailto:datamanagement@geomar.de).

[View Terms of Use](#)

[Login for unrestricted access...](#)

## Latest Legs/Expeditions

Label	Departure - Return	Chief-Scientist
ISL_00612_1	2012/09/20 - 2012/09/21	Vieira, Nuno
M81/2B	2010/03/29 - 2010/04/21	Werner, Reinhard
POS463	2013/11/16 - 2013/12/22	Bialas, Jörg

[View more...](#)

## Latest Models

Label	Created	Responsible Person
MPI-ESM	2012/09/27	Wittke, Franziska
MPI-ESM	2012/09/13	Mohr, Viktoria
piControl-ocean	2012/09/12	Marun, Thomas

[View more...](#)

## Latest Links

Link	Created
<a href="#">Print Publication linked to Leg M77/4</a>	2012/09/27
<a href="#">Print Publication linked to Leg SO218</a>	2012/09/27
<a href="#">Print Publication linked to Leg SO218</a>	2012/09/27

[View more...](#)

**Search for Events/Locations...**

...by Year :  -

...by Latitude [°]:  -

...by Longitude [°]:  -

2D-Multi-Channel-Seismic | 2D-MCS [ ]  
3D Full Tensor Gradiometry | 3-D FTG  
3D-Multi Channel Seismic | 3D-MCS [P-Cable]  
ARGO Float | ARGOF  
Aanderaa  
Access Point | AP  
Acoustic Doppler Current Profiler | ADCP  
Acoustic Modem | AM

...by Gear:

...by Leg/Expedition:

...by Locality:

...on sea: ☒ ...on land: ☐

...with files: ☐

[Go...](#)

Source: Kiel Data Management Team, <https://portal.geomar.de/web/guest/kdmi>

# GEOMAR – OceanRep

The screenshot displays the GEOMAR OceanRep website. The header features the GEOMAR logo and the text 'Helmholtz Centre for Ocean Research Kiel'. Below the header is a navigation bar with tabs: INSTITUTE, RESEARCH, STUDY, DISCOVER, and SERVICE. The main content area is titled 'OceanRep' and shows search results for the query 'toste'. The results are displayed in a list format, with each entry including the authors, title, journal, and a small thumbnail image. The left sidebar contains a 'QUICK SEARCH' box and a 'BROWSE' section with links to various categories like 'Author', 'Research division', and 'Document type'. The right sidebar has a 'Login' button and an 'RSS' feed icon.

**GEOMAR** Helmholtz Centre for Ocean Research Kiel

**OceanRep**

INSTITUTE RESEARCH STUDY DISCOVER SERVICE

**OceanRep**

> OceanRep Home  
> Contact

QUICK SEARCH

Simple Search  
Advanced Search

BROWSE

Author  
Research division  
Document type  
Year  
Course of Study

LATEST

Peer-reviewed Articles  
All

ABOUT US  
GEOMAR LIBRARY  
OPEN ACCESS  
POLICIES  
STATEMENTS  
HFI P

All fields matches "toste"

Displaying results 1 to 7 of 7.  
[Refine search](#) | [New search](#)

Order the results: by year (most recent first)

Export 7 results as: ASCII Citation

Export

No shelf selected [change](#)

1. Banyte, D., Tanhua, T., Visbeck, M., Wallace, D. W. R., Karstensen, J., Krahmann, G., Schneider, A., Stramma, L. and Dengler, M. (2012) *Diapycnal diffusivity at the upper boundary of the tropical North Atlantic oxygen minimum zone* Journal of Geophysical Research - Oceans, 117 (C9). C09016. DOI [10.1029/2011JC007762](#). Item availability may be restricted.

2. Körtzinger, A., Bange, H., Breitbarth, E., Marandino, C., Quack, B. and Tanhua, T. (2012) *Ocean-Atmosphere Coupling – Biogeochemical Processes and Material Exchange: Contributions from the Department of Chemical Oceanography* [Poster] In: Helmholtz-Centre for Ocean Research Kiel (GEOMAR): Scientific Advisory Board Evaluation, 09.05.2012, Kiel, Germany . Item not available from this repository.

3. Stramma, L., Czeschel, R., Visbeck, M. and Tanhua, T. (2012) *Stagnant flow and eddies in the oxygen minimum zone south of the Cape Verde Islands* [Talk] In: Tropical Atlantic Variability meeting, 12.09.2012, Kiel, Germany . Item not available from this repository.

4. Tanhua, T. and Keeling, R. F. (Submitted) *Changes in column inventories of carbon and oxygen in the Atlantic Ocean* Biogeosciences Discussions, 9 (7). pp. 8039-8073. DOI [10.5194/bgd-9-8039-2012](#).

5. Tanhua, T. (2011) *1. Wochenbericht M84/3* UNSPECIFIED.

6. Tanhua, T. (2011) *2. Wochenbericht M84/3* UNSPECIFIED.

# GEOMAR – OceanRep link to data

The screenshot displays the GEOMAR website interface. The top navigation bar includes links for 'Startseite', 'English', 'Suche', 'Projekte, Publikationen', and 'Mitarbeiter'. The main header features the GEOMAR logo and the text 'Helmholtz-Zentrum für Ozeanforschung Kiel'.

The left sidebar contains a navigation menu with sections like 'INSTITUT', 'FORSCHEN', and 'STUDIUM'. Under 'FORSCHEN', there is a section for 'OceanRep' with links to 'OceanRep Startseite', 'Kontakt', and 'SCHNELLSUCHE'. Below this, there are links for 'Einfache Suche', 'Erweiterte Suche', 'BLÄTTERN', 'Autor', 'Forschungsbereich', 'Publikationsart', 'Jahr', and 'Studiengang'. At the bottom of the sidebar, there is a section for 'NEUZUGÄNGE' with links to 'Artikel – begutachtet' and 'Alle'.

The main content area shows a search result for 'FS Poseidon Fahrtbericht / Cruise Report POS408/1, 13.01.-02.03.2011'. The author is 'Schmidt, Mark, Devey, Colin und Eisenhuth, Ralf'. The report is a 'Text' document, and a 'Download (14MB)' link is provided. The official URL is 'http://eprints.ifm-geomar.de/13289/'.

The 'Supplementary Data' section includes an 'Abstract' and a 'Typ des Eintrags' (Type of Entry) section. The 'Abstract' describes the Jeddah Transect Project, a multi-disciplinary cruise offshore Saudi Arabia (Red Sea). The 'Typ des Eintrags' section lists various fields of study: Petrology, Meeresgeologie, Geodynamics, Sedimentology, Oceanography, Microbiology, Marine Biology, Red Sea, hydrothermalism, hydrocarbon formation, brine biogeochemistry, ocean in statu nascendi, bathymetry, plankton, food chain of pelagic organisms.

The 'Forschungsbereich' (Research Area) section lists several projects: 'OceanRep > GEOMAR > FB2 Marine Biogeochemie > FB2-MG Marine Geosysteme', 'OceanRep > GEOMAR > FB4 Dynamik des Ozeanbodens > FB4-MUHS Magmatische und hydrothermale Systeme', and 'DOI etc.: 10.13289/IFM-GEOMAR\_Rep\_46\_2011'. A list of 'Expeditionen/Modelle' (Expeditions/Models) is also provided, including 'Cruise POS408', 'Leg POS408/1', 'Leg POS408/2', and 'Leg POS408/3'. The 'Leg POS408/1' link is circled in orange.

The right sidebar shows a detailed view of the 'Leg POS408/1' data. It includes a 'Context' dropdown menu set to 'List all...'. Below this, there are tabs for 'General Leg Info', 'Events (68)', 'Files (3)', and 'Related Links (2)'. The 'General Leg Info' tab is active, showing details about the leg: 'Leg/Phase: POS408/1', 'Cruise/Expedition: POS408', 'Platform: Poseidon (POS)', and 'Departure/Return: Departure: 2011/01/13 - Jeddah (Saudi Arabia) Return: 2011/01/31 - Jeddah (Saudi Arabia)'. There is also an 'Add Web Infos' button.

# GEOMAR – OceanRep link to Pangaea

The screenshot displays the GEOMAR website interface. At the top, the GEOMAR logo and the text 'Helmholtz Centre for Ocean Research Kiel' are visible. Below this is a navigation bar with tabs for 'INSTITUTE', 'RESEARCH', 'STUDY', 'DISCOVER', and 'SERVICE'. The 'OceanRep' section is active, showing a sidebar with search options and a main content area. The main content area features a record titled 'Enhanced modern heat Transfer to the Arctic by warm Atlantic Water' by Spielhagen, Robert F., Werner, Kirstin, Sørensen, Steffen Aagaard, Zamelczyk, Katarzyna, Kandiano, Evgeniya S., Budéus, Gereon, Husum, Katrine, Marchitto, Thomas M., and Hald, Morten. The record is linked to a PANGAEA entry. A red box highlights the PANGAEA record details, including the citation, abstract, and project information. A red arrow points from the 'Download (910Kb)' link in the OceanRep record to the PANGAEA record. A red circle highlights the 'Supplementary data' link in the OceanRep record, which also points to the PANGAEA record. The PANGAEA record includes a map of the Arctic region and a list of keywords.

**GEOMAR** Helmholtz Centre for Ocean Research Kiel

**OceanRep**

> OceanRep Home  
> Contact  
QUICK SEARCH  
Simple Search  
Advanced Search  
BROWSE  
Author  
Research division  
Document type  
Year  
Course of Study  
LATEST

**Enhanced modern heat Transfer to the Arctic by warm Atlantic Water**  
Spielhagen, Robert F., Werner, Kirstin, Sørensen, Steffen Aagaard, Zamelczyk, Katarzyna, Kandiano, Evgeniya S., Budéus, Gereon, Husum, Katrine, Marchitto, Thomas M., and Hald, Morten (2011). pp. 450-453. DOI [10.1126/science.1197397](https://doi.org/10.1126/science.1197397)  
2011\_Science\_Spielhagen\_et\_al\_450-3.pdf - Restricted to Registered users only  
[Download \(910Kb\)](#) | [Contact](#)  
Official URL: <http://dx.doi.org/10.1126/science.1197397>  
Supplementary data: [DATA](#)  
Abstract  
The Arctic is responding more rapidly to global warming than most other areas on our planet. Northward flowing Atlantic Water is the major means of heat advection towards the Arctic and strongly affects the sea ice distribution. Records of its natural variability are critical for the understanding of feedback mechanisms and the future of the Arctic climate system, but continuous historical records reach back only ~150 years. Here, we present a multidecadal scale record of ocean temperature variations during the last 2000 years, derived from marine sediments off Western Svalbard (79°N). We find that early-21st-century temperatures of Atlantic Water entering the Arctic Ocean are unprecedented over the past 2000 years and are presumably linked to the Arctic Amplification of global warming.

**PANGAEA®**  
Data Publisher for Earth & Environmental Science

**Data Description**  
Citation: Spielhagen, RF et al. (2011): Planktic foraminiferal distribution and stable isotope ratios of sediment core MSM05/5\_712-1 from the Arctic Ocean. doi:10.1594/PANGAEA.755114, Supplement to: Spielhagen, Robert F.; Werner, Kirstin; Sørensen, Steffen Aagaard; Zamelczyk, Katarzyna; Kandiano, Evgeniya S.; Budéus, Gereon; Husum, Katrine; Marchitto, Thomas M.; Hald, Morten (2011): Enhanced modern heat transfer to the Arctic by warm Atlantic water. *Science*, 331(6016), 450-453, doi:10.1126/science.1197397  
Abstract: The Arctic is responding more rapidly to global warming than most other areas on our planet. Northward flowing Atlantic Water is the major means of heat advection towards the Arctic and strongly affects the sea ice distribution. Records of its natural variability are critical for the understanding of feedback mechanisms and the future of the Arctic climate system, but continuous historical records reach back only ~150 years. Here, we present a multidecadal scale record of ocean temperature variations during the last 2000 years, derived from marine sediments off Western Svalbard (79°N). We find that early-21st-century temperatures of Atlantic Water entering the Arctic Ocean are unprecedented over the past 2000 years and are presumably linked to the Arctic Amplification of global warming.  
Project(s): Integrierte Analyse zwischeneiszeitlicher Klimadynamik (INTERDYNAMIK)  
Coverage: Median Latitude: 78.914531 \* Median Longitude: 6.773434 \* South-bound Latitude: 78.911000 \* West-bound

Not logged in (log in or sign up)  
Always quote citation when using data!  
Show Map Google Earth RIS BioTeX  
Hybrid  
Greenland Sea  
Barents Sea  
Imagery ©2012 NASA, TerraMetrics - Terms of Use

Kielprints is a similar service for Kiel at large:

<http://eprints.uni-kiel.de>

Recently, a collaboration with the Eprints Group at the University of Southampton started.

# EPrints 3.3 for Research Data

## Data & Projects Demonstrator

[Home](#)
[About](#)
[Browse by Year](#)
[Browse by Publications by Type](#)
[Browse by Non-publications by Type](#)

Logged in as Dr Leslie Carr | [Manage deposits](#) | [Manage records](#) | [Profile](#) | [Saved searches](#) | [Review](#) | [Admin](#) | [Logout](#)

### N,N'-bis(2-Bromophenyl)urea

Huth, Susanne L. and Threlfall, Terry L. and Hursthouse, Michael B. (2008) *N,N'-bis(2-Bromophenyl)urea*. [Experiment]

**Abstract**

InChI=1/C13H10Br2N2O/c14-9-5-1-3-7-11(9)16-13(18)17-12-8-4-2-6-10(12)15/h1-8,16-17H/q-1 Data collection parameters: Chemical formula: C13 H10 Br2 N2 O; Crystal morphology: Needle; Crystal system: Orthorhombic; Space group symbol: P 21 21 2; Cell length a: 11.5691(4); Cell length b: 11.5772(4); Cell length c: 4.66520(10); Cell angle alpha: 90.00; Cell angle beta: 90.00; Cell angle gamma: 90.00; Data collection temperature: 120(2); Refinement results: Solution figure of merit: 0.0221; R Factor (Obs): 0.0275; R Factor (All): 0.0295; Weighted R Factor (Obs): 0.0588; Weighted R Factor (All): 0.0597;

Raw/Original Stage	Processed/Analysed Stage	Final/Results Stage
 <b>Data - Image (JPEG)</b> 2008ish144_hk0.jpg <a href="#">Download (61Kb)</a>   <a href="#">Preview</a> HK1 cross-section	 <b>Data - Plain Text</b> 2008ish144.prp <a href="#">Download (6Kb)</a>	 <b>Data - Plain Text</b> 2008ish144.cif <a href="#">Download (13Kb)</a>
 <b>Data - Image (JPEG)</b> 2008ish144_h0.jpg <a href="#">Download (45Kb)</a>   <a href="#">Preview</a> HDL cross-section	 <b>Data - Plain Text</b> 2008ish144.res <a href="#">Download (4Kb)</a>	 <b>Data - HTML</b> 2008ish144.cml <a href="#">Download (3639b)</a>
 <b>Data - Image (JPEG)</b> 2008ish144_okl.jpg <a href="#">Download (31Kb)</a>   <a href="#">Preview</a> OKL cross-section	 <b>Data - Plain Text</b> 2008ish144_xs.lst <a href="#">Download (33Kb)</a>	 <b>Data - Plain Text</b> 2008ish144.tcf <a href="#">Download (69Kb)</a>
 <b>Data - HTML</b> 2008ish144.htm <a href="#">Download (9Kb)</a>	 <b>Data - Plain Text</b> 2008ish144_xl.lst <a href="#">Download (39Kb)</a>	
 <b>Data - Plain Text</b> 2008ish144.hkl <a href="#">Download (235Kb)</a>		

### Miscellaneous Stage

 <b>Data - Image (JPEG)</b> 2008ish144_h1l.jpg <a href="#">Download (49Kb)</a>   <a href="#">Preview</a> HKL cross-section	 <b>Data - Image (JPEG)</b> 2008ish144_1kl.jpg <a href="#">Download (38Kb)</a>   <a href="#">Preview</a> 1KL cross-section
 <b>Data - Image (JPEG)</b> 2008ish144_hk1.jpg <a href="#">Download (64Kb)</a>   <a href="#">Preview</a> HK1 cross-section	 <b>Data - Image (GIF)</b> 2008ish144_ellipsoid.gif <a href="#">Download (21Kb)</a>   <a href="#">Preview</a>
 <b>Data - Plain Text</b> 2008ish144.p4p <a href="#">Download (101Kb)</a>	 <b>Data - Plain Text</b> 2008ish144.mol <a href="#">Download (2073b)</a>
 <b>Data - Plain Text</b> 2008ish144.ins <a href="#">Download (4Kb)</a>	 <b>Data - Plain Text</b> 2008ish144.pdb <a href="#">Download (4Kb)</a>
 <b>Data - Plain Text</b> 2008ish144.inchi <a href="#">Download (958b)</a>	

### Other Documents

 <b>README file - HTML</b> plan.html <a href="#">Download (22Kb)</a>
 <b>README file - RDF/XML</b> plan.rdf <a href="#">Download (9Kb)</a>
 <b>Data - Generic Spreadsheet (Tab Separated)</b> <a href="#">Download (514b)</a> Summary metadata taken from original

## Data & Projects Demonstrator

[Home](#)
[About](#)
[Browse by Year](#)
[Browse by Subject](#)
[Browse by Division](#)
[Browse by Author](#)

## Workflow (bioinformatics)

Data & Projects Demonstrator is powered by [EPrints 3](#) which is developed by the [School of Electronics and Computer Science](#) at the University of

Experiment (crystallography)

Source: Lesly Carr,  
University of Southampton 21

# Agenda

- What's the problem / challenge that I'm talking about?
  - And what I'm not talking about today ...
- What's the current state?
  - And what you could already do ...
- **What to expect in the future?**
  - **And what you could do ...**

# PubFlow



## Toward Publication Workflows

### Funded:

Software Engineering Group, University Kiel

### Associated:

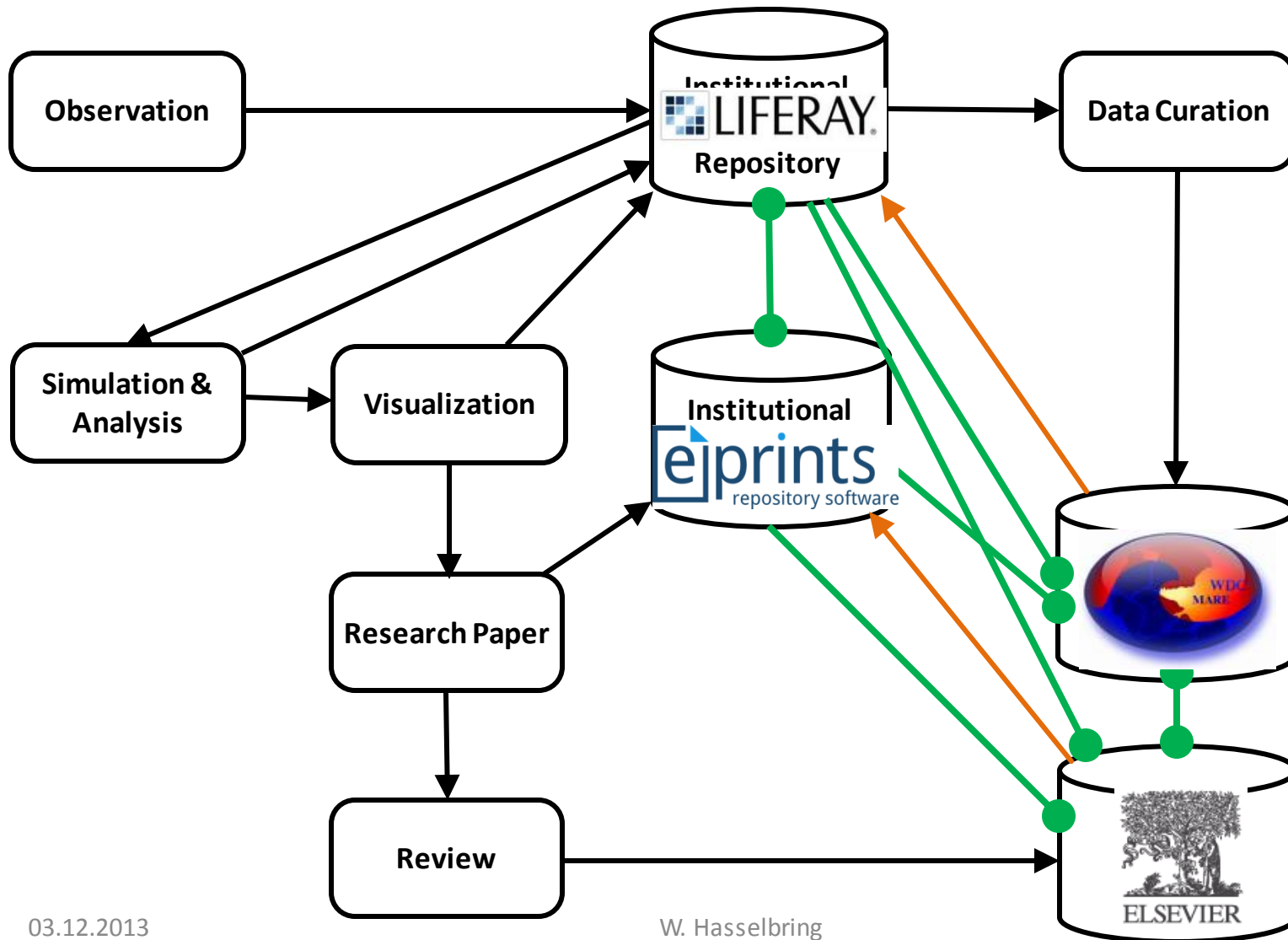
- Excellence cluster “Future Ocean”
- Data and computing center of GEOMAR
- **Library** of GEOMAR
- Computing center of University Kiel
- **Library** of University Kiel
- ZBW  
German National **Library** of Economics -  
Leibniz Information Centre for Economics



<http://www.pubflow.uni-kiel.de/>



# Data and Paper Flow (in Ocean Science)





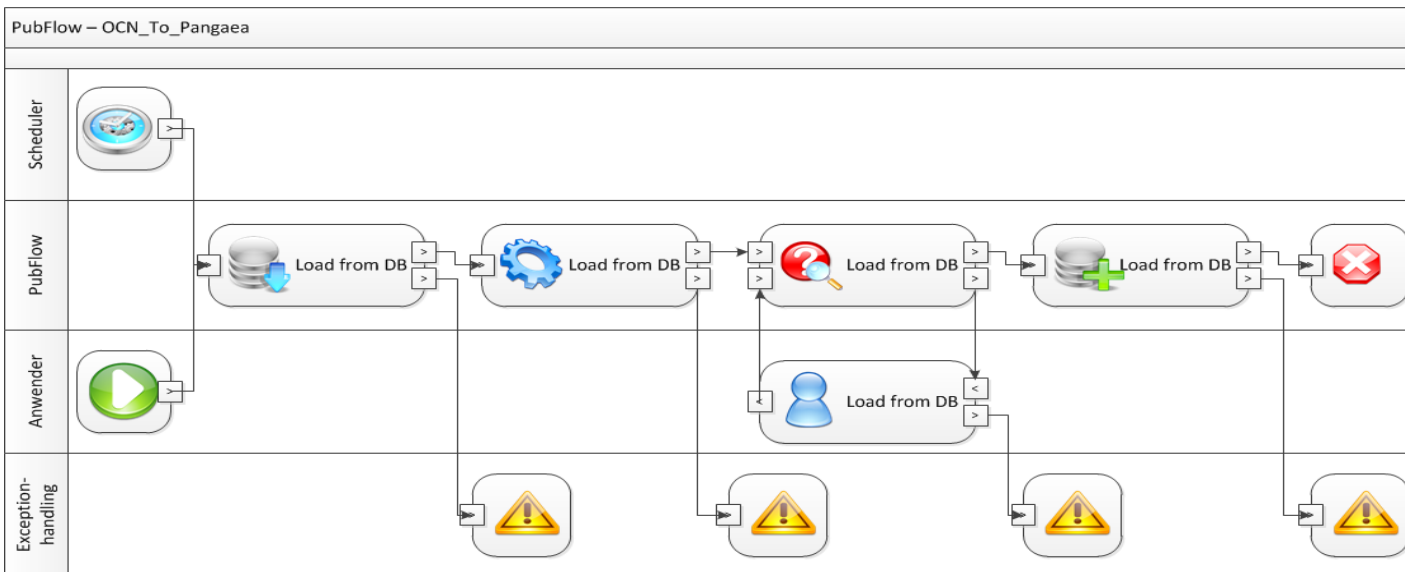
# The initial PubFlow case study

## CTD Rosette System

(Conductivity /  
Temperature /  
Depth/pressure  
sensor).



# CTD Workflow



run process

2

Log Id	7001
PID	PID
Login	LOGIN
Source	SOURCE
Author	AUTHOR
Project	PROJECT
Topology	TOPOLOGY
status	STATUS
Zipfile (z.B. /home/test/...)	/home/test/7001.txt
Reference	REFERENCE
FileName	FILENAME
Comment	COMMENT

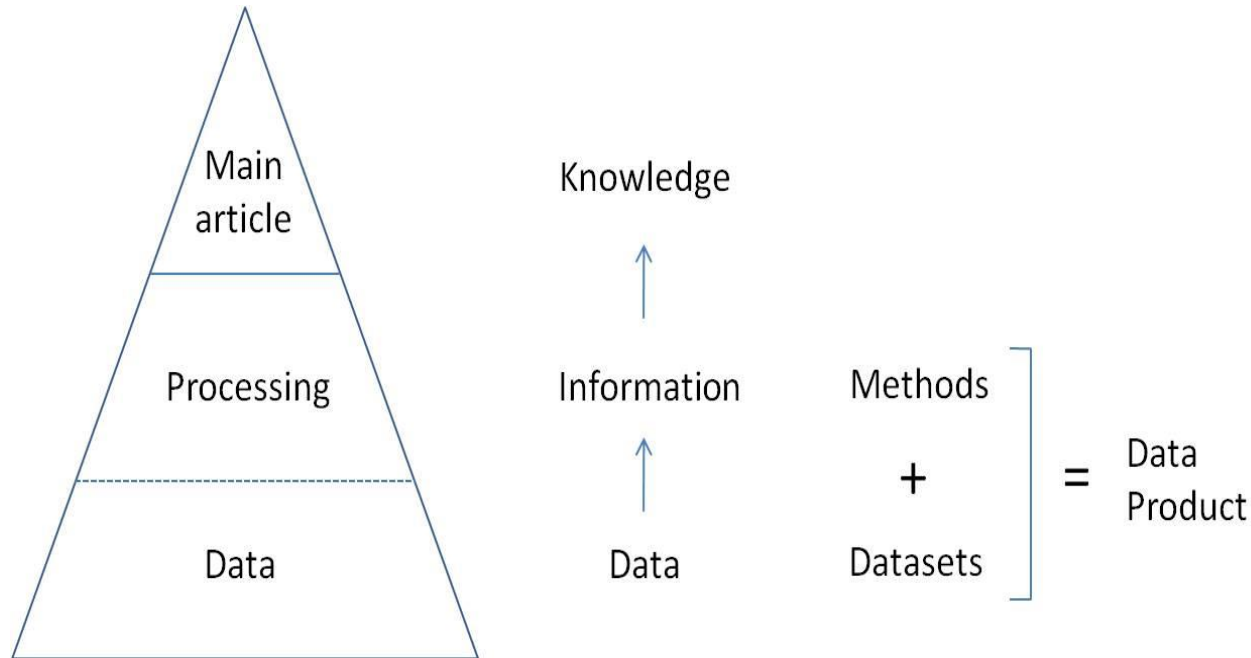
Do it! close



# Automatically collecting provenance data in PubFlow

- Provenance information describes the origins and the history of (research) data in its life cycle, and the process by which it arrived:
  - A description of all modifications and changes
  - A list of all methods applied
  - A list of all systems and persons involved
  - Answers to
    - when,
    - where,
    - how,
    - why.
- Given the documented history of an object, the object attains an **authority** that allows scholars to appreciate its importance, whereas, in the absence of such history, the object may be treated with some skepticism.
- Usually collected and used in databases, data warehouses, workflow systems, scientific data processing, Web services, etc.
- We employ Kieker [Rohr et al. 2008, van Hoorn et al. 2009] for automatically collecting provenance data [Brauer and Hasselbring 2012, 2013a, 2013b].

# Need to save data + processing (not yet addressed in PubFlow)



*Algorithms + Data Structures = Programs*

Source: Kunze, John A; Cruse, Patricia; Hu, Rachael; Abrams, Stephen; Hastings, Kirk; Mitchell, Catherine; et al. (2011). Practices, Trends, and Recommendations in Technical Appendix Usage for Selected Data-Intensive Disciplines. <http://escholarship.org/uc/item/9jw4964t>

# The case for open computer programs

Darrel C. Ince<sup>1</sup>, Leslie Hatton<sup>2</sup> & John Graham-Cumming<sup>3</sup>

- “We argue that, with some exceptions, anything less than the release of source programs is intolerable for results that depend on computation.
- The vagaries of hardware, software and natural language will always ensure that exact reproducibility remains uncertain,
  - but withholding code increases the chances that efforts to reproduce results will fail.”



# SPEC Research Group

Wednesday, 02 January 2013

- Home
- News
- FAQ
- Mission and Charter
- Working Groups
- ICPE Proceedings
- Benchmarking Portal
- Awards
- Tools**
- Submission
- Newsletter
- Member Organizations
- Steering Committee
- Contact

## Quantitative Evaluation and Analysis Tools

This page hosts a repository of peer-reviewed tools for quantitative system evaluation and analysis. The published tools have undergone a thorough review process by multiple independent experts to ensure high quality and relevance to the community. The review process covers important quality factors, including maturity, availability and usability.

## Overview of Hosted Tools

Tool name	Description	License	Requirements
<a href="#">Fabian</a>	Fabian is a framework and tool to automate the creation and running of multi-tier server performance tests/workloads.	CDDL 1.0	Unix or Windows. Java 5 or later.
<a href="#">Kieker</a>	Kieker is a framework for application performance monitoring and dynamic software analysis.	Apache License, Version 2.0	Java 5 or later.

## Important Links

- [RG Cloud Working Group](#)
- [IDS Benchmarking Working Group](#)
- [SPEC Distinguished Dissertation Award](#)
- [Quantitative Evaluation and Analysis Tools](#)
- [Latest Newsletter Issue](#)
- [ICPE International Conference](#)

See, for instance: [Waller and Hasselbring 2012]

**Source** code and **benchmark** code / data are available at  
<http://kieker-monitoring.net/>

# Executable, data papers



More info at

<http://www.executablepapers.com/>

<http://www.articleofthefuture.com/>

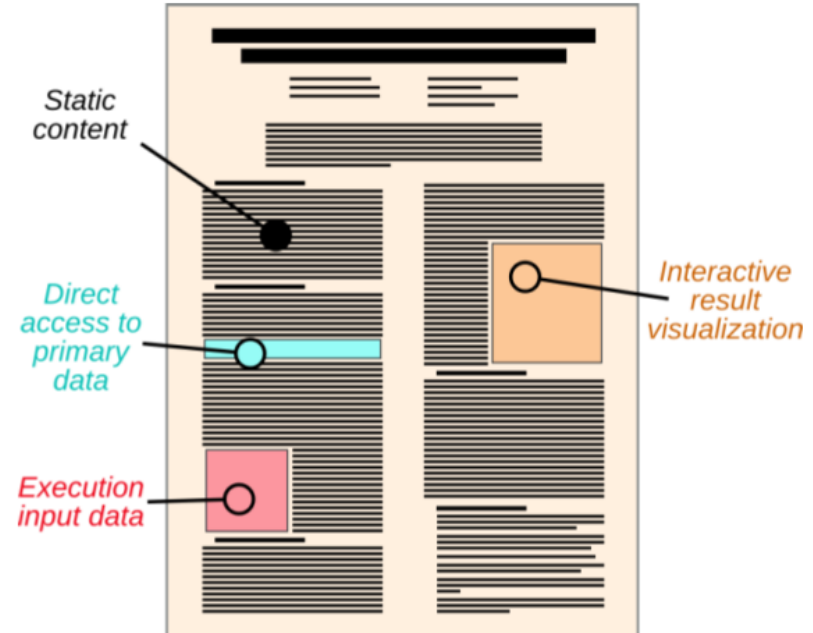
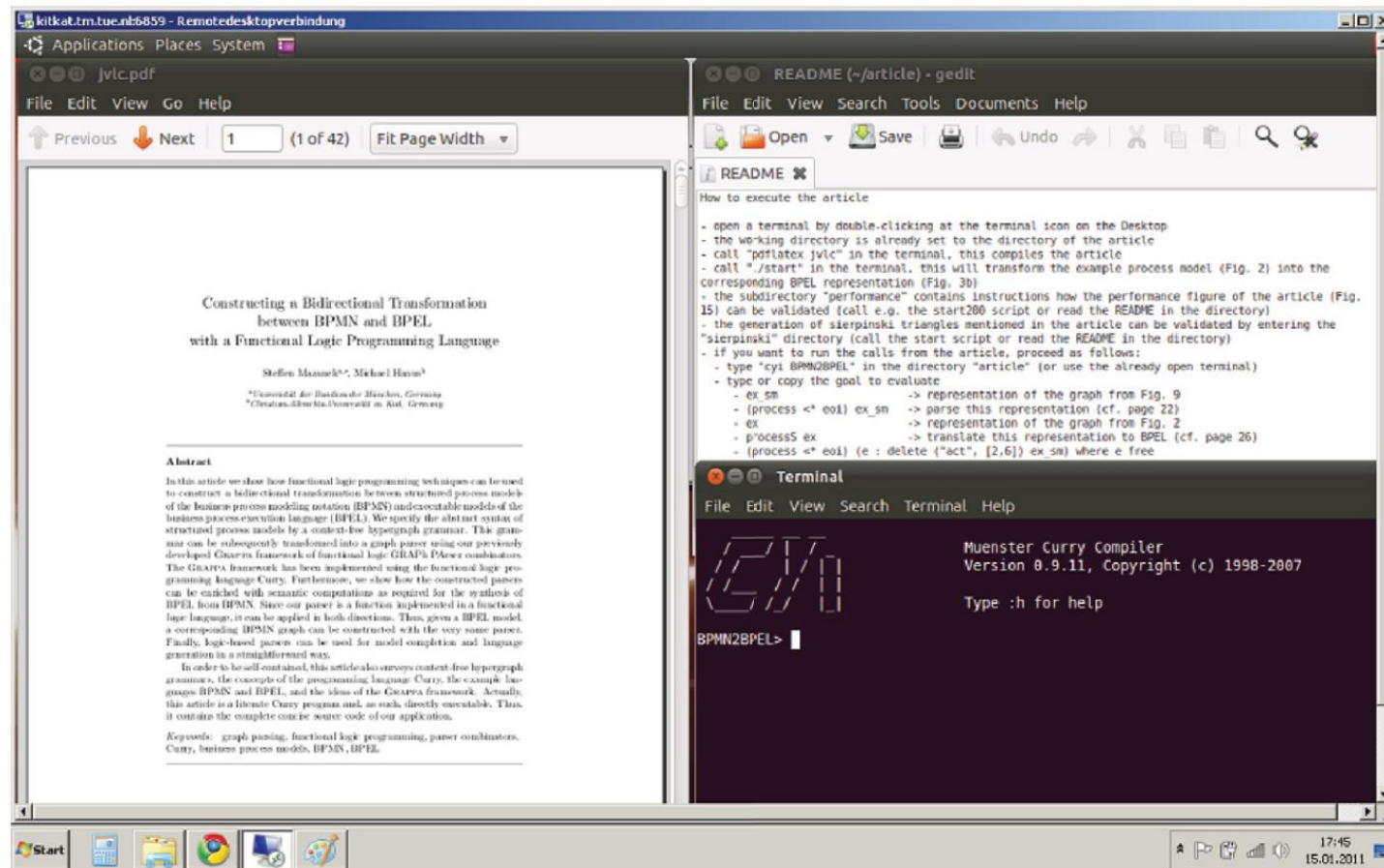


Fig. 1: Conceptual view of the executable paper. Static content (the body of the publication) is extended by interactive access primary data and reenact computations in order to validate the presented conclusions or navigate result spaces. For peer review approval, readers can also obtain access to the underlying code of the experiments presented in the publication. The i and can be integrated with the Publisher's portal.



# An Example Executable Paper



**SHARE: a web portal for creating and sharing executable research papers**

<http://sites.google.com/site/executablepaper/>

<http://dx.doi.org/10.1016/j.procs.2011.04.062>

<http://www.springerlink.com/content/y1488178640l6412/>



# What about social networks?



## Social Networking Meets Software Development:

Perspectives from GitHub,  
MSDN, Stack Exchange,  
and TopCoder

Specific social networks for academics exist such as ResearchGate (<http://www.researchgate.net>) or Mendeley (<http://www.mendeley.com/>).

Andrew Begel, Microsoft

Jan Bosch, Chalmers University of Technology

Margaret-Anne Storey, University of Victoria

IEEE Software 30(1): 26-28, 2013

Digital Object Identifier: 10.1109/MS.2013.13\_

# Policies and Incentives

- Funding agencies, such as the DFG, require strategies for research data management
  - Institutional data policies and infrastructures may help
  - “Modular” data management policy for Kiel Marine Sciences may already be reused
- Published data and code may be listed in CVs
- Cost benefit analysis of the DRYAD repository
  - Papers with published data receive higher citation counts:
    - Piwowar, Vision, Whitlock: “Data archiving is a good investment”, Nature 473(285), 2011 <http://dx.doi.org/10.1038/473285a>

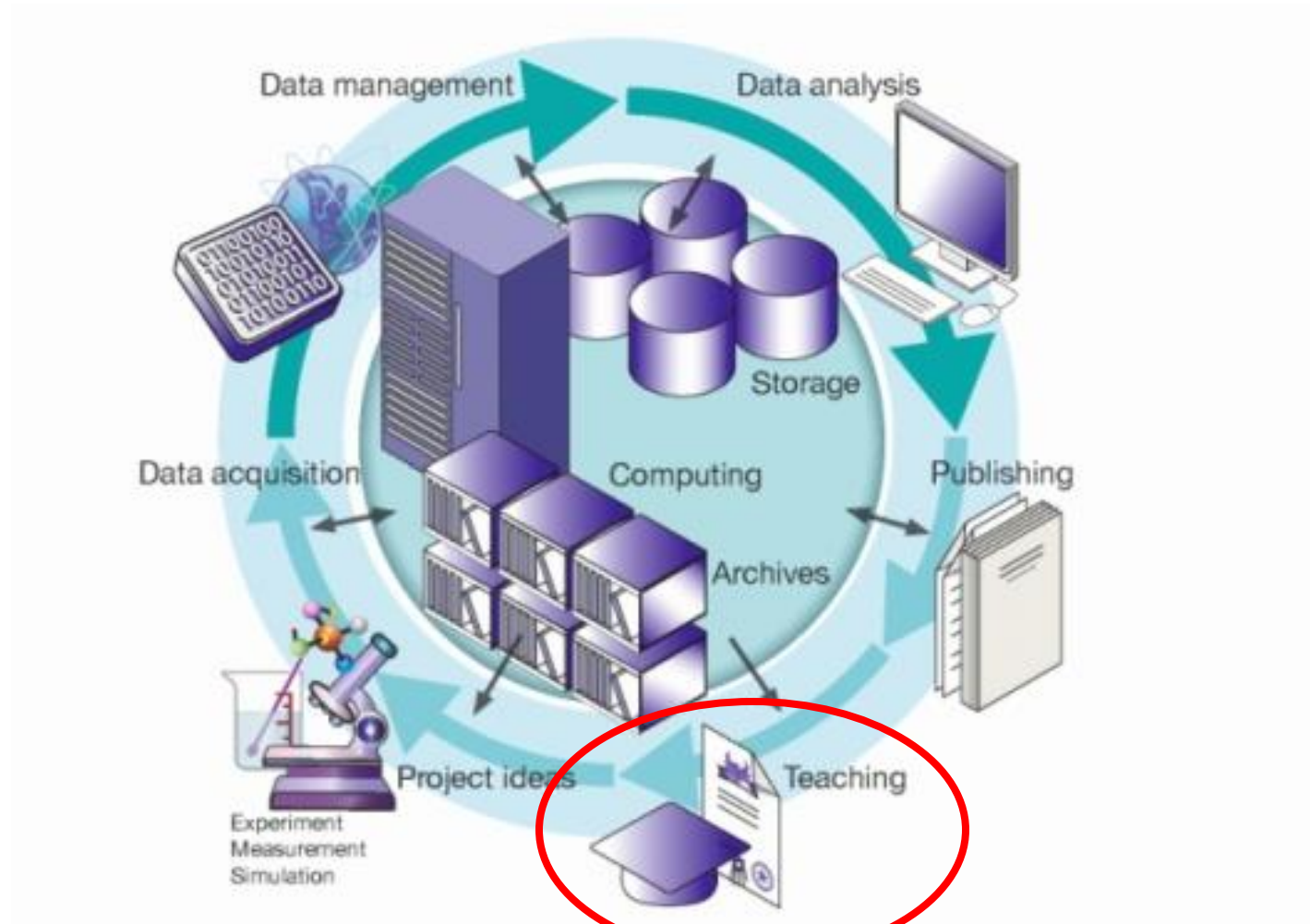
# Metrics scenarios for the data publishing models (and comparison with the current scientific publication model)

Types of metrics	Currently available tools with possibilities for “data metrics”	Metric dimensions	Models			
			Scientific publication	Data publication	Data Publication	
					Stand-alone data publications	Journal data publications
Data publication & citation-based metrics	<ul style="list-style-type: none"> <li>- Data Citation Index (Web of Science)</li> <li>- Google Scholar</li> <li>- Scopus</li> <li>- Microsoft Academic Search</li> <li>- DataCite</li> </ul>	Size-dependent	Yes	Difficult (1)	Yes (4)	Yes
		Size-independent				
		- Direct average performance	Yes	No	Yes	Yes
		- Source-based performance	Yes	No	Yes (3)	Yes
Altmetrics-based metrics	<ul style="list-style-type: none"> <li>- ImpactStory</li> <li>- Twitter, Facebook</li> </ul>	Social media indicators	Yes	Yes	Yes	Yes
	<ul style="list-style-type: none"> <li>- Mendeley</li> <li>- CiteULike</li> </ul>	Readership counts	Yes	No	Yes	Yes
	<ul style="list-style-type: none"> <li>- Repositories</li> <li>- Data Journals</li> </ul>	Downloads & views counts (DUI metrics)	Yes	Difficult (2)	Yes	Yes

<http://www.knowledge-exchange.info/datametrics>

# The data life cycle

(<http://www.scc.kit.edu/forschung/l sdf.php>)



# Summary

- If you are only interested in getting a Ph.D.,
  - this talk was not really of interest to you, sorry.
- If you are (also) interested in scientific impact, publish
  - research papers,
  - research data,
  - documented code, and
  - do networking with related stakeholders.
- Software systems and services may help

# References

- [Brauer and Hasselbring 2012] Brauer, P. C. and Hasselbring, W. “Capturing provenance information with a workflow monitoring extension for the Kieker framework.” In: The Third International Workshop on the role of Semantic Web in Provenance Management , 28.5.2012, Heraklion, Kreta. <http://eprints.uni-kiel.de/19636/>
- [Brauer and Hasselbring 2013a] Brauer, P. C. and Hasselbring, W. “PubFlow: a scientific data publication framework for marine science.” In: IMDIS 2013 International Conference on Marine Data and Information Systems, 23.-25.09.2013, Lucca, Italy. <http://eprints.uni-kiel.de/22399/>
- [Brauer and Hasselbring 2013b] Brauer, P. C. and Hasselbring, W. (2013) “PubFlow: provenance-aware workflows for research data publication.” In: 5th USENIX Workshop on the Theory and Practice of Provenance (TaPP '13), April 2-3, 2013, Lombard. <http://eprints.uni-kiel.de/21112/>
- [Rohr et al. 2008] M. Rohr, A. van Hoorn, J. Matevska, N. Sommer, L. Stoeber, S. Giesecke, and W. Hasselbring. “Kieker: Continuous monitoring and on demand visualization of Java software behavior.” In C. Pahl, editor, Proceedings of the IASTED International Conference on Software Engineering 2008 (SE'08), pages 80–85, Feb. 2008. ISBN 978-0-88986-715-4. <http://eprints.uni-kiel.de/14496/>
- [van Hoorn et al. 2009] A. van Hoorn, M. Rohr, W. Hasselbring, J. Waller, J. Ehlers, S. Frey, and D. Kieselhorst. Continuous monitoring of software services: Design and application of the Kieker framework.” Technical Report TR-0921, Department of Computer Science, University of Kiel, Germany, Nov. 2009b. <http://eprints.uni-kiel.de/14459/>
- [van Hoorn et al. 2012] A. van Hoorn, J. Waller, and W. Hasselbring. Kieker: A framework for application performance monitoring and dynamic software analysis. In Proceedings of the 3rd ACM/SPEC International Conference on Performance Engineering (ICPE 2012), pages 247–248. ACM, Apr. 2012. ISBN 978-1-4503-1202-8. <http://eprints.uni-kiel.de/14418/>
- [Waller and Hasselbring 2012] J. Waller and W. Hasselbring: “A Comparison of the Influence of Different Multi-core Processors on the Runtime Overhead for Application-Level Monitoring.” Springer LNCS vol. 7303, 2012, <http://eprints.uni-kiel.de/15425/>