

# Good Scientific Data Management Practice

Prof. Dr. Wilhelm (Willi) Hasselbring  
Software Engineering Group

<http://se.informatik.uni-kiel.de/>

ISOS, March 22th, 2017



# 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 ...**

# Scientific misconduct and other challenges

- There exist several examples of scientific misconduct, such as the case Jan Hendrik Schön.
  - I 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 former 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  - Sources for calibration data are named (some are unpublished) but again we cannot reconstruct the specific input data and parameters used.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.

# Reproducibility to Rectify Errors

PNAS 2016 vol. 113 no. 28 7900–7905, DOI: 10.1073/pnas.1602413113

## Cluster failure: Why fMRI inferences for spatial extent have inflated false-positive rates

Anders Eklund<sup>a,b,c,1</sup>, Thomas E. Nichols<sup>d,e</sup>, and Hans Knutsson<sup>a,c</sup>

### Significance

Functional MRI (fMRI) is 25 years old, yet surprisingly its most common statistical methods have not been validated using real data. Here, we used resting-state fMRI data from 499 healthy controls to conduct 3 million task group analyses. Using this null data with different experimental designs, we estimate the incidence of significant results. In theory, we should find 5% false positives (for a significance threshold of 5%), but instead we found that the most common software packages for fMRI analysis (SPM, FSL, AFNI) can result in false-positive rates of up to 70%. These results question the validity of some 40,000 fMRI studies and may have a large impact on the interpretation of neuroimaging results.

# Gene name errors are widespread in the scientific literature

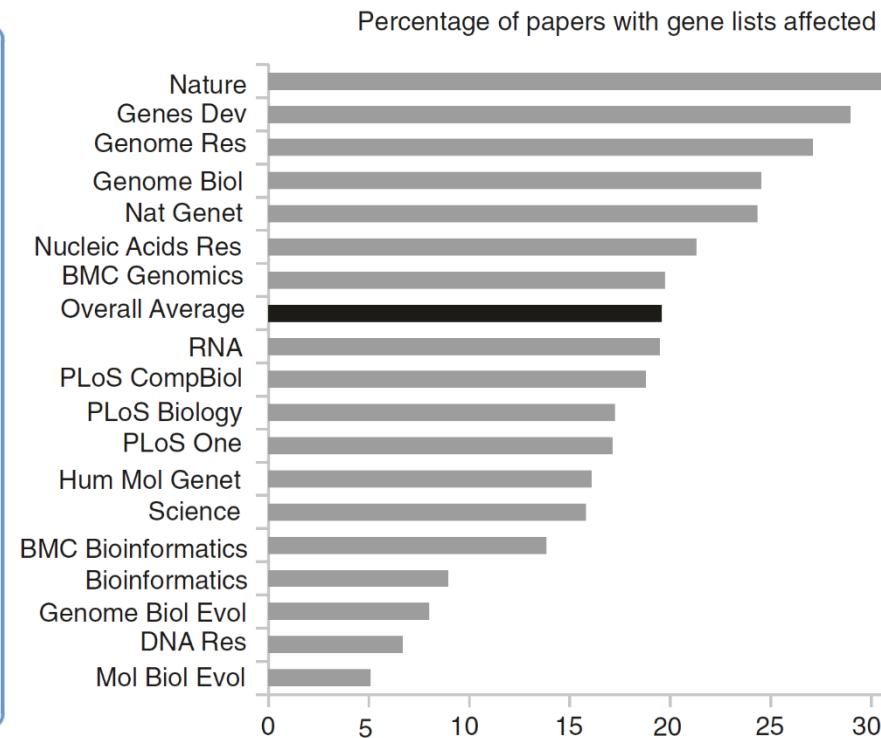
Mark Ziemann<sup>1</sup>, Yotam Eren<sup>1,2</sup> and Assam El-Osta<sup>1,3\*</sup>

## Abstract

The spreadsheet software Microsoft Excel, when used with default settings, is known to convert gene names to dates and floating-point numbers. A programmatic scan of leading genomics journals reveals that approximately one-fifth of papers with supplementary Excel gene lists contain erroneous gene name conversions.

**Keywords:** Microsoft Excel, Gene symbol, Supplementary data

**Abbreviations:** GEO, Gene Expression Omnibus; JIF, journal impact factor



# 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://doi.org/10.1002/9783527679188.oth1>)

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

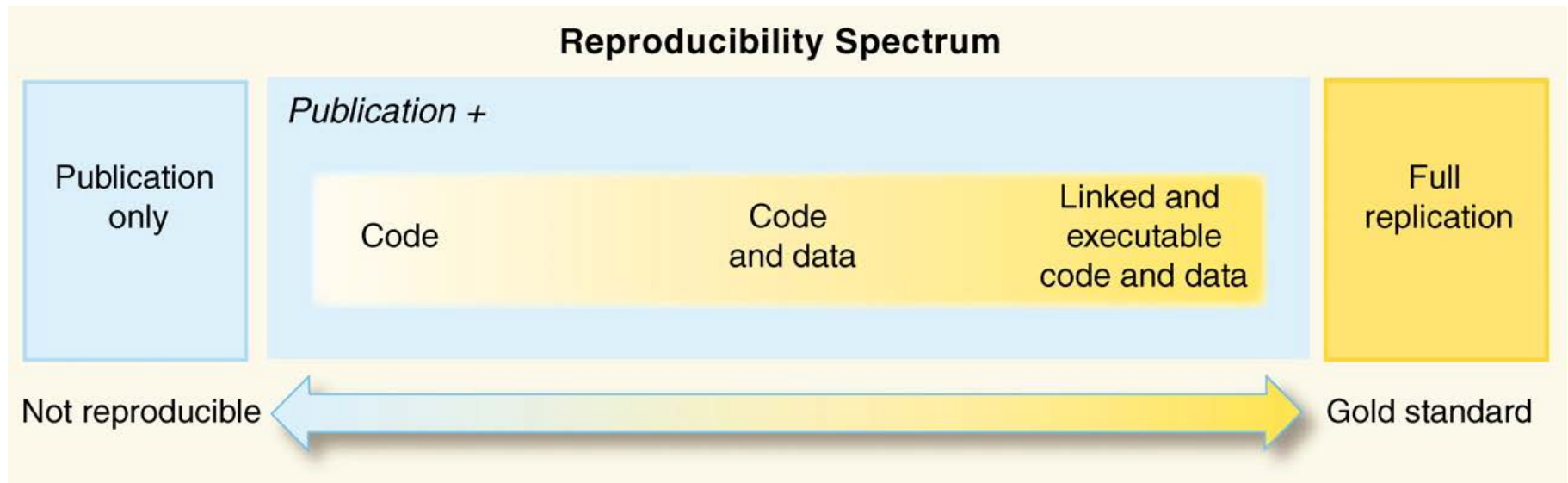
Isaac Newton, 1676

PERSPECTIVE

## Reproducible Research in Computational Science

Roger D. Peng

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





# 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://plagiarism-detector.com/>

<https://www.plagaware.com/>



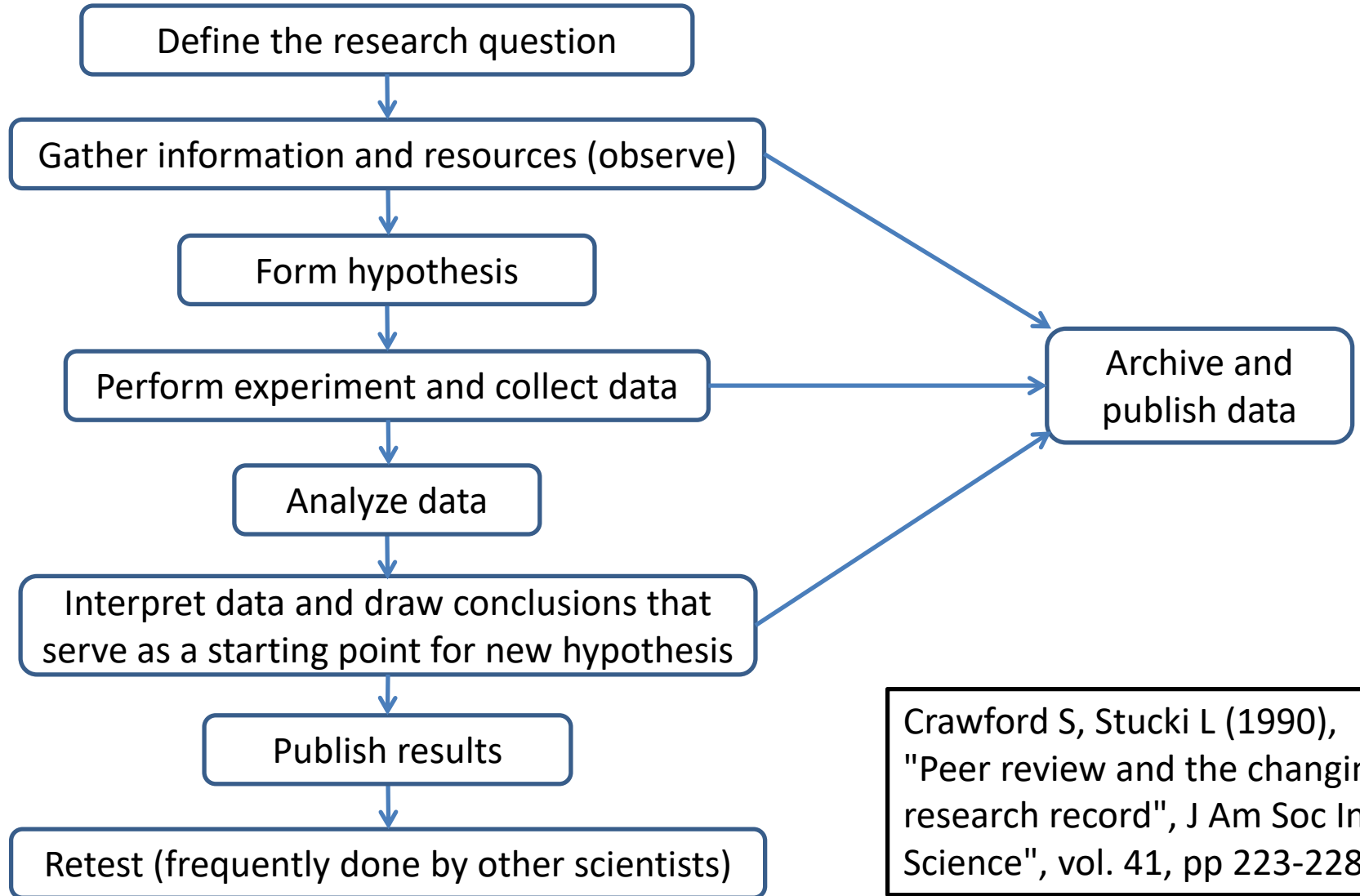
# What I'm not talking about?

- Establishing Software Engineering best practices in computational science, such as
  - **Version and configuration management**  
[Ploski et al. 2007]
  - **Quality management**  
[Waller et al. 2015, Hasselbring & Steinacker 2017]
  - **Software architecture design and modeling** [Hasselbring 2002, Hasselbring 2006, Reussner & Hasselbring 2008]
  - **Domain-specific programming languages** [Johanson & Hasselbring 2016, Johanson et al. 2016b, Johanson et al. 2017a]
  - **Parallel and distributed programming**  
[Hasselbring 1994, Hasselbring 2000, Wulf et al. 2016]
- 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 ...

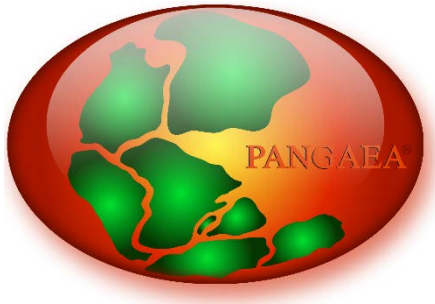
# Research Workflows



Crawford S, Stucki L (1990), "Peer review and the changing research record", J Am Soc Info Science", vol. 41, pp 223-228

# Data Repositories (Services): Examples

[Registry: <http://www.re3data.org/>]



**PANGAEA®**

**Data Publisher for Earth & Environmental Science**

<https://www.pangaea.de/>



**World Data Center  
for Climate  
Hamburg**



<https://www.dkrz.de/daten/wdcc/>



<http://zenodo.org/>

Projekt

Workshops

**Forschungsdatenmanagement**

Aktivitäten

Kolloquium  
Forschungsdatenmanagement

Umfrage zum Umgang mit  
Forschungsdaten

Kontext und Informationen

Plattform Data Management

Partner

Team

## Forschungsdatenmanagement

Forschungsdaten sind die Daten, die in einer Virtuellen Forschungsumgebung vorgehalten werden, und sind die Grundlage wissenschaftlichen Arbeitens.

Allgemeine Informationen zum Forschungsdatenmanagement und Informationen mit Bezug zum Forschungsstandort Kiel sind auf den folgenden Seiten zusammengestellt.

### Aktivitäten

Lokale Angebote, und Information zu Vernetzungen und Kooperationen über den Standort Kiel hinaus sind [hier gelistet](#).

### Kolloquium Forschungsdatenmanagement

Das Kolloquium Forschungsdatenmanagement ist eine regelmäßig stattfindende Veranstaltungsreihe, in der die Möglichkeit zum Austausch über den Umgang mit Forschungsdaten über die Grenzen der verschiedenen Disziplinen an der CAU und ein Forum zur Diskussion aktueller Themen geboten werden. [Hier weiterlesen...](#)

### Umfrage zum Umgang mit Forschungsdaten

Zur Ermittlung der vielfältigen Anforderungen der verschiedenen Fachdisziplinen wurde eine eine Umfrage zum Thema 'Umgang mit Forschungsdaten am Forschungsstandort Kiel' durchgeführt. [Hier weiterlesen...](#)

### Kontext und Information

Nationale, europäische und internationale Rahmenbedingungen und damit zusammenhängende Informationen zum Thema Forschungsdaten sind [hier gelistet](#).

### Plattform Data Management

Die Plattform Data Management bündelt die Kompetenzen bereits existierender Datenmanagement-Initiativen der Forschungsschwerpunkte Kiel Marine Science, Kiel Life Science und SECC mit dem Ziel, ein einrichtungübergreifendes Forschungsdatenmanagement für den Forschungsstandort Kiel zu entwickeln und etablieren.

Neben der [Kiel Data Management Infrastructure \(KDMI\)](#) für Meereswissenschaften bilden die [DFG-Projekte 'PubFlow'](#) der AG Software Engineering und ['VFU@Kiel'](#) des Servicezentrums [Forschung und Innovation](#) und Rechenzentrums die wesentliche Grundlage für flexibles Forschungsdatenmanagement an der CAU.

▼ Ziele

▼ Aufgaben

▼ Leitlinie zum Umgang mit Forschungsdaten

▼ AG Forschungsdatenmanagement

▼ Kolloquium Forschungsdatenmanagement

▼ Projekte

 Kontakt

→ Thilo Paul-Stüve

☎ 880-1969

✉ [tpaul-stueve@uv.uni-kiel.de](mailto:tpaul-stueve@uv.uni-kiel.de)

 Umfeld

→ AG Datenmanagement

→ Exzellenzcluster "Ozean der Zukunft"

→ Exzellenzcluster "Entzündungen an Grenzflächen"

→ KDMI – Kiel Data Management Infrastruktur

→ PubFlow – Publication Workflows for Scientific Data

 Deutsche Forschungsgemeinschaft

<http://www.uni-kiel.de/vfu/de/forschungsdatenmanagement>


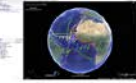







# Kiel Data Management Portal



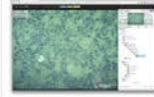




Refer to the "hands on" seminars

<https://portal.geomar.de/>

## "Open Access" to Research Data

	<ul style="list-style-type: none"><li>■ <b>OSIS- Kiel (Ocean Science Information System):</b> Allows data exchange and description for expeditions, numeric models and experiments. Expeditions are linked to KML-files (Google-Earth compatible), allowing a visualization of all stations of a cruise expedition.<ul style="list-style-type: none"><li>■ How To - <a href="#">OSIS Manual (PDF)</a></li><li>■ How To - <a href="#">Integrate your Data</a> (GEOMAR and research projects only)</li><li>■ Go to <a href="#">OSIS</a></li><li>■ Contact: <a href="#">Data Management Team</a></li></ul></li></ul>
	<ul style="list-style-type: none"><li>■ <b>KDMI Geo-Search:</b> Search cruise tracks, print publications, sampling locations, glider tracks, sample descriptions and detailed cruise information as georeferenced information in GoogleEarth.<ul style="list-style-type: none"><li>■ Go to <a href="#">KDMI GEO-Search</a></li><li>■ Contact: <a href="#">Data Management Team</a></li></ul></li></ul>
	<ul style="list-style-type: none"><li>■ <b>GEOMAR OPeNDAP (THREDDS):</b> The OPeNDAP server allows access to published data of model experiments. The data sets are interlinked in OceanRep and OSIS.<ul style="list-style-type: none"><li>■ How To - <a href="#">Integrate your Data</a> (GEOMAR and research projects only)</li><li>■ Go to <a href="#">GEOMAR OPeNDAP (THREDDS)</a></li><li>■ Contact: <a href="#">Data Management Team</a></li></ul></li></ul>
	<ul style="list-style-type: none"><li>■ <b>Boknis Eck Database:</b> The Boknis Eck Time Series Station is located at the entrance of the Eckernförde Bay (54°31.2' N, 10°02.5' E) in the southwestern Baltic Sea. Monthly sampling at Boknis Eck began on 30 April 1957. Thus, Boknis Eck is one of the oldest - still operated - time series sites worldwide. Registered users can access the database (biogeochemical data).<ul style="list-style-type: none"><li>■ Go to <a href="#">Boknis Eck Database web page</a></li><li>■ Contact: <a href="#">Prof. Dr. Hermann W. Bange</a></li></ul></li></ul>
	<ul style="list-style-type: none"><li>■ <b>OceanRep GEOMAR:</b> OceanRep is an open access digital collection containing the research output of GEOMAR staff and students. Included are journal articles, conference papers, book chapters, theses and more - with fulltext, if available. OceanRep is managed by the GEOMAR Library and is fully supported by the Data Management Team.<ul style="list-style-type: none"><li>■ Go to <a href="#">OceanRep GEOMAR web page</a></li><li>■ Contact: <a href="mailto:Bibliothek.bibliothek@geomar.de">Bibliothek.bibliothek@geomar.de</a></li></ul></li></ul>
	<ul style="list-style-type: none"><li>■ <b>PANGAEA:</b> Open Access and citable datasets related to GEOMAR with persistent identifiers in a long-term archive.<ul style="list-style-type: none"><li>■ How To - <a href="#">PANGAEA Data Submission Manual (PDF)</a></li><li>■ <a href="#">Submit your Data</a> (GEOMAR and research projects only)</li><li>■ Go to <a href="#">PANGAEA web page &amp; WIKI</a></li><li>■ Contact: <a href="#">Data Management Team</a></li></ul></li></ul>
	<ul style="list-style-type: none"><li>■ <b>Data Portal German Marine Research (MaNIDA):</b> The data portal offers an integrative framework for coherent discovery, visualization, download and dissemination of scientific data originating from nationally operated research platforms and monitoring facilities (vessels, observatories, campaigns, web services). The content originates from German marine research institutions ranging from cruise-related metadata to archived and near real time data, data products, reports and publications.<ul style="list-style-type: none"><li>■ Go to <a href="#">MaNIDA web page</a></li><li>■ Contact: <a href="#">Data Management Team</a></li></ul></li></ul>

## "Access Controlled" Research Data (ongoing research)

	<ul style="list-style-type: none"><li>■ <b>DSHIP land system:</b> Contains tracks, stations and underway measurements from cruises by German research vessels.<ul style="list-style-type: none"><li>■ How To - <a href="#">DSHIP Manual (PDF)</a></li><li>■ Go to <a href="#">DSHIP web page</a></li><li>■ Contact: <a href="#">Data Management Team</a></li></ul></li></ul>
	<ul style="list-style-type: none"><li>■ <b>Media Server - (ProxSys):</b> The media server stores videos and images with metadata description and provides access to larger collections for image reuse in other applications such as DIAS or <a href="#">GEOMAR image database</a>.<ul style="list-style-type: none"><li>■ How To - <a href="#">ProxSys Manual (PDF)</a></li><li>■ How To - <a href="#">Integrate your Data</a> (GEOMAR and research projects only)</li><li>■ Go to <a href="#">Media Server web page</a></li><li>■ Contact: <a href="#">Data Management Team</a></li></ul></li></ul>
	<ul style="list-style-type: none"><li>■ <b>DIAS:</b> A web based image annotation system under current development for marine research.<ul style="list-style-type: none"><li>■ Go to <a href="#">DIAS web page</a></li><li>■ Contact: <a href="#">Dr. Timm Schöning</a></li></ul></li></ul>
	<ul style="list-style-type: none"><li>■ Version Control: <a href="#">Subversion &amp; GitLab</a> (Version Control System for source code -models, scripts, matlab...-).<ul style="list-style-type: none"><li>■ How To - <a href="#">Subversion Manual</a> and <a href="#">GitLab Manual</a></li><li>■ How To - <a href="#">Integrate your Data</a> (GEOMAR and research projects only)</li><li>■ Go to <a href="#">Subversion web page</a> or <a href="#">GitLab web page</a></li><li>■ Contact: <a href="#">Data Management Team</a></li></ul></li></ul>
	<ul style="list-style-type: none"><li>■ <b>MEMENTO (Marine Methane an Nitrous Oxide):</b> MEMENTO aims to become a valuable tool for identifying regions of the world ocean that should be targeted in future work to improve the quality of air-sea flux estimates.<ul style="list-style-type: none"><li>■ Go to <a href="#">MEMENTO web page</a></li><li>■ Contact: <a href="#">Dr. Annette Kock</a></li></ul></li></ul>
	<ul style="list-style-type: none"><li>■ <b>CVOO database:</b> The database for CTD bottle data allows search and export of environmental data from selected cruises in the context of a research project.<ul style="list-style-type: none"><li>■ Go to <a href="#">CVOO database</a></li><li>■ Contact: <a href="#">Dr. Björn Fiedler</a></li></ul></li></ul>
	<ul style="list-style-type: none"><li>■ <b>Cyclone database:</b> Cyclone track database of the Northern hemisphere derived from ERA/Interim or NCEP reanalyses. Tracks of a certain area can be searched, exported and displayed in Google Earth.<ul style="list-style-type: none"><li>■ Go to <a href="#">Cyclone database</a></li><li>■ Contact: <a href="#">Dr. Andreas Lehmann</a></li></ul></li></ul>



# Kiel Data Management Infrastructure – OSIS: Ocean Science Information System

[Overview](#)
[Expeditions](#)
[Numerical Models](#)
[Experiments](#)
[More...](#)
[Login...](#)

Context:

## Kiel Ocean Science Information System for Expeditions, Numeric Models, 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 [AT] geomar.de.

### View Terms of Use

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

...by Year :  -

...by Latitude [°]:  -

...by Longitude [°]:  -

...by Gear:

2D-Multi-Channel-Seismic |  2D-MCS [ ]  
 3D Full Tensor Gradiometry |  3-D FTG  
 3D-Multi Channel Seismic |  3D-MCS [P-Cable]  
 80 Feet Fishing Net |  80-Feet-Net  
 ADCP-Mooring |  TRBM-ADCP  
 ARGO Float |  ARGOFL  
 Aanderaa  
 Access Point | AP

...by Leg/Expedition:

...by Locality:

...on sea:  ...on land:

...with files:

### Last changed Expeditions

Label	Departure - Return	Chief-Scientist
AL502	2017/09/13 - 2017/09/26	Hübscher, Christian
AL499	2017/08/14 - 2017/08/26	Süling, Jörg
AL492	2017/05/01 - 2017/05/10	Schwarzer, Klaus

### Last changed Models

Label	Updated	Responsible Person
UVic-MOBI-limWC D=32,epsBD=2	2016/08/15	Somes, Christopher J.
PISCES-ORCA2-H0 T2	2016/08/15	Segsneider, Joachim
PISCES-ORCA2-H6 A	2016/08/15	Segsneider, Joachim

### Last changed Experiments

Experiment	Updated	Responsible Person
Wind Daten 2002-2015	2016/11/17	van der Lee, Eefke
test	2016/11/15	van der Lee, Eefke
KOSMOS 2010 Svalbard	2016/08/05	Riebesell, Ulf

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

# OceanRep link to OSIS data, and back

## OceanRep

> OceanRep Startseite

> Kontakt

SCHNELLSUCHE

> Einfache Suche

> Erweiterte Suche

BLÄTTERN

> Autor

> Forschungsbereich

> Publikationsart

> Jahr

> Studiengang

NEUZUGÄNGE

> Artikel – begutachtet

> Alle

Legs » Cruise POS494 » Leg POS494

General Leg Info Events (23) Files (2) Related Links (3)

...for Cruise (0) ...for this Leg (3) ...for Events (0)

Create new related link for Leg POS494...

Type of Link	Link Address (Path) / Description	Date created	Action...
DSHIP	<a href="http://dship.geomar.de">http://dship.geomar.de</a> DSHIP Sensor data available at GEOMAR	2016-03-22 09:56	Action...
Print Publication	<a href="http://oceanrep.geomar.de/35024">http://oceanrep.geomar.de/35024</a> Poseidon No. 494, leg 1 [POS494/1] Technical Cruise Report, 04th Feb. - 07th Feb., 2016, Las Palmas - La Estaca, El Hierro (Spain)	2016-12-05 11:27	Action...
URL	<a href="http://www.eskp.de/mit-tauchboot-jago-zum-unterwasservulkan">http://www.eskp.de/mit-tauchboot-jago-zum-unterwasservulkan</a> Earth System Knowledge Platform: Reading the Landscape	2016-01-29 14:20	Action...



Available under License Creative Commons: Attribution 4.0.

[Download \(3536Kb\)](#) | [Vorschau](#)

Typ des Eintrags: Berichte (Fahrtbericht)

Schlagwörter: RV Poseidon, POS494/1, Technical Cruise Report, El Hierro, OBS

Forschungsbereich: [OceanRep](#) > [GEOMAR](#) > [FB4 Dynamik des Ozeanbodens](#) > [FB4-GDY Marine Geodynamik](#)

■ [Leg POS494](#)

Expeditionen/Modelle/Experimente:

Hinterlegungsdatum: 05 Dez 2016 10:32

Letzte Änderung: 05 Dez 2016 10:32

URI: <http://oceanrep.geomar.de/id/eprint/35024>

# OceanRep link to Pangaea



## OceanRep

> OceanRep Startseite

> Kontakt

SCHNELLSUCHE

> Einfache Suche

> Erweiterte Suche

BLÄTTERN

> Autor

> Forschungsbereich

> Publikationsart

> Jahr

> Studiengang

NEUZUGÄNGE

> Artikel – begutachtet

> Alle

ÜBER UNS

GEOMAR BIBLIOTHEK

OPEN ACCESS

POLICIES

GRUNDSÄTZE

HILFE

FAQS

STATISTIK

Anfahrt | Kontakt | Sitemap | Impressum

## It is the economy, stupid! Projecting the fate of fish populations using ecological-economic modeling

**Quaas, Martin, Reusch, Thorsten B.H., Schmidt, Jörn, Tahvonen, Olli und Voss, Rudi (2016) *It is the economy, stupid! Projecting the fate of fish populations using ecological-economic modeling* Global Change Biology, 22 (1), pp. 264-270. DOI [10.1111/gcb.13060](https://doi.org/10.1111/gcb.13060).**



Text  
Quaas\_et\_al-2016-Global\_Change\_Biology.pdf - publizierte Version  
Restricted to Nur registrierte Benutzer  
[Download \(334Kb\)](#) | [Kontakt](#)



Text (Figure S1. Past development (according to ICES/CCAT stock assessments) in filled circles, sample of stochastic future development (empty circles), and most likely future development for a scenario with economic parameters set constant at 2008 levels,...)

[gcb13060-sup-0001-FigureS1.pdf](#) - Begleitmaterial

Restricted to Nur registrierte Benutzer  
[Download \(41Kb\)](#) | [Kontakt](#)



Text (Figure S2. Past development (according to ICES/CCAT stock assessments) in filled circles, sample of stochastic future development (empty circles), and most likely future development for the baseline scenario of future spawning stock sizes, as in Fig. 1..)

[gcb13060-sup-0002-FigureS2.pdf](#) - Begleitmaterial

Restricted to Nur registrierte Benutzer  
[Download \(42Kb\)](#) | [Kontakt](#)



Text (Figure S3. Past development (according to ICES/CCAT stock assessments) in filled circles, and model runs starting in 1988 with sample of stochastic development (empty circles), and most likely development for the baseline scenario)

[gcb13060-sup-0003-FigureS3.pdf](#) - Begleitmaterial

Restricted to Nur registrierte Benutzer  
[Download \(33Kb\)](#) | [Kontakt](#)



Text (Data S1. Materials and methods, detailing the bio-economic modeling approach. Fishing cost functions are parametrized based on literature data where available (Armason et al., 2004; Diekert et al., 2010; Froese & Quaas, 2011, 2012) or otherwise using ...)

[gcb13060-sup-0004-DataS1.doc](#) - Begleitmaterial

Restricted to Nur registrierte Benutzer  
[Download \(984Kb\)](#) | [Kontakt](#)

Supplementary Data:



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

SEARCH SUBMIT ABOUT CONTACT

**Citation:** **Quaas, Martin; Reusch, Thorsten BH; Schmidt, Jörn O; Tahvonen, Olli; Voss, Rüdiger (2016):** Projecting the fate of fish populations using ecological-economic modeling. doi:10.1594/PANGAEA.856741, **Supplement to:** Quaas, M et al. (2016): It is the economy, stupid! Projecting the fate of fish populations using ecological-economic modeling. *Global Change Biology*, **22(1)**, 264-270, doi:10.1111/gcb.13060

Always quote above citation when using data! You can download the citation in several formats below.

[RIS Citation](#) [BibTeX Citation](#) [Text Citation](#) [Facebook](#) [Twitter](#) [Google+](#)

**Abstract:** Four marine fish species are among the most important on the world market: cod, salmon, tuna, and sea bass. While the supply of North American and European markets for two of these species - Atlantic salmon and European sea bass - mainly comes from fish farming, Atlantic cod and tunas are mainly caught from wild stocks. We address the question what will be the status of these wild stocks in the mid-term future, in the year 2048, to be specific. Whereas the effects of climate change and ecological driving forces on fish stocks have already gained much attention, our prime interest is in studying the effects of changing economic drivers, as well as the impact of variable management effectiveness. Using a process-based ecological-economic multispecies optimization model, we assess the future stock status under different scenarios of change. We simulate (i) technological progress in fishing, (ii) increasing demand for fish, and (iii) increasing supply of farmed fish, as well as the interplay of these driving forces under different scenarios of (limited) fishery management effectiveness. We find that economic change has a substantial effect on fish populations. Increasing aquaculture production can dampen the fishing pressure on wild stocks, but this effect is likely to be overwhelmed by increasing demand and technological progress, both increasing fishing pressure. The only solution to avoid collapse of the majority of stocks is institutional change to improve management effectiveness significantly above the current state. We conclude that full recognition of economic drivers of change will be needed to successfully develop an integrated ecosystem management and to sustain the wild fish stocks until 2048 and beyond.

**Project(s):** [Biological Impacts of Ocean Acidification \(BIOACID\)](#)

**License:** Creative Commons Attribution 3.0 Unported

**Size:** 2 datasets

**Download Data**  
[Download ZIP file containing all datasets as tab-delimited text \(use the following character encoding: UTF-8: Unicode \(PANGAEA default\)\)](#)

**Datasets listed in this Collection**

1. **Quaas, M (2016):** Annual growth rate of fish populations using ecological-economic modeling (Fig. 3). doi:10.1594/PANGAEA.856740
2. **Quaas, M (2016):** Spawning stock biomass of fish populations using ecological-economic modeling (Fig.1 and 2). doi:10.1594/PANGAEA.856739

Kielprints is a similar service for Kiel at large:  
<http://eprints.uni-kiel.de>

# 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



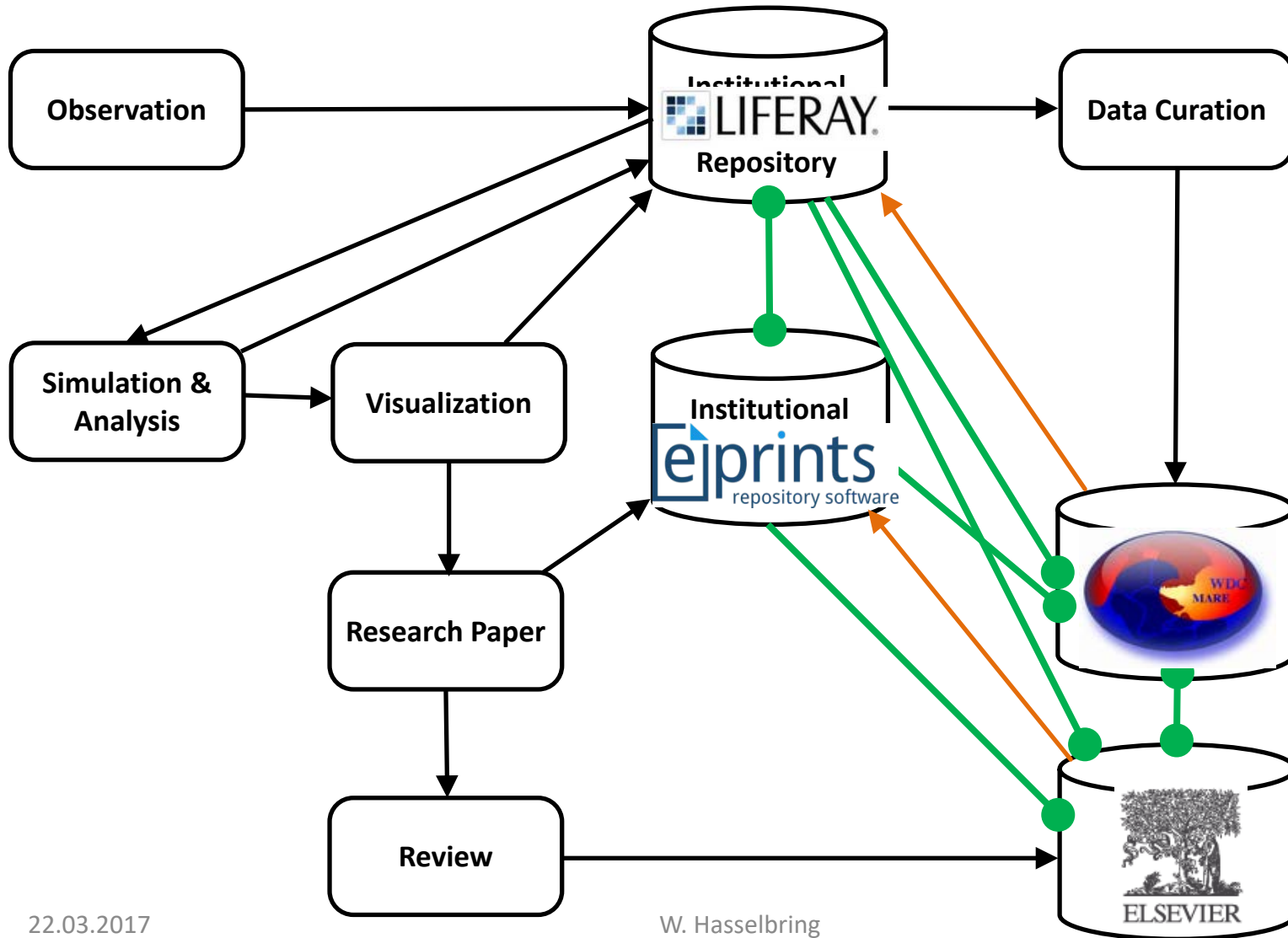
### Associated:

- 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

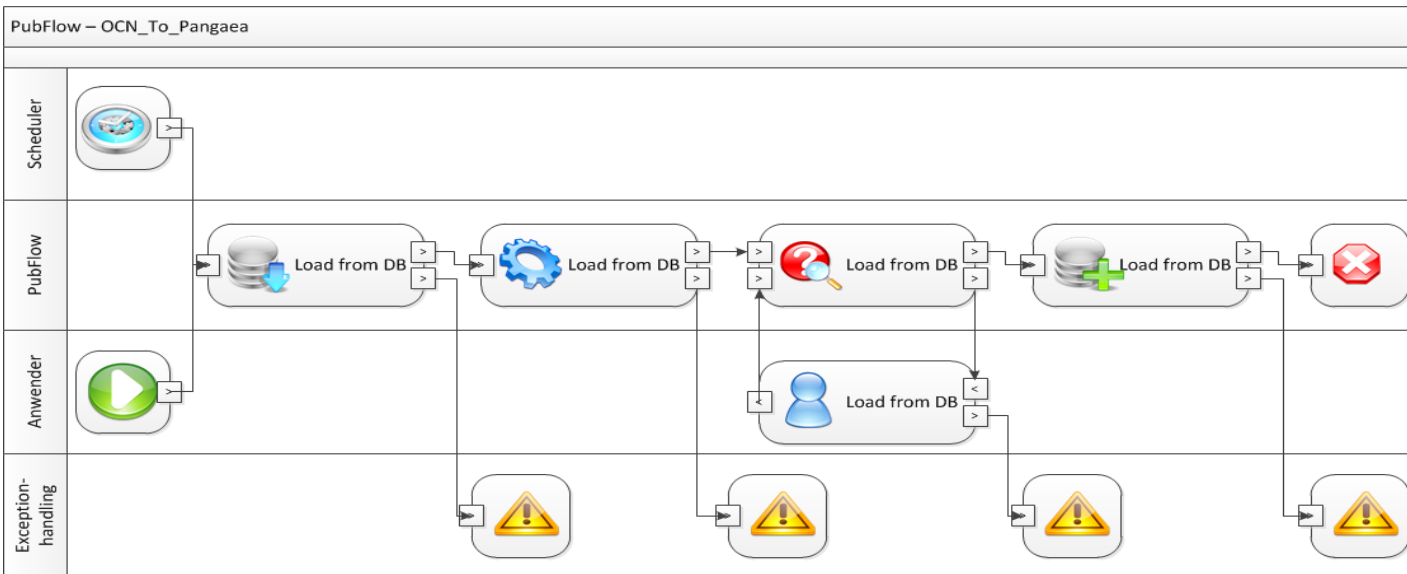


<http://www.pubflow.uni-kiel.de/> [Brauer & Hasselbring 2013]

# Data and Paper Flow (in Ocean Science)



# CTD Workflow



run process

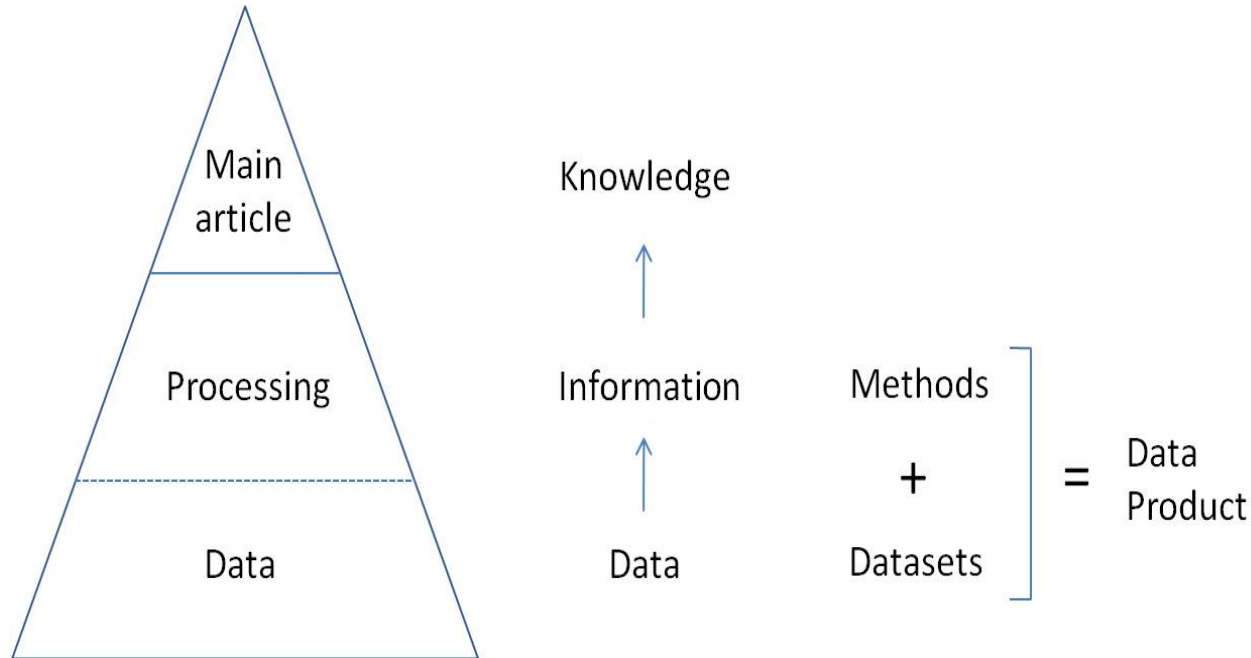
2

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

Do it! close



# 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.”

# Viewpoint

## The Real Software Crisis: Repeatability as a Core Value

*Sharing experiences running artifact evaluation committees for five major conferences.*

“Science advances faster when we can build on existing results, and when new ideas can easily be measured against the state of the art.”

*Repeatability, not necessarily reproducibility*

Several ACM SIGMOD, SIGPLAN, and SIGSOFT conferences have initiated artifact evaluation processes.

# What are we doing?

## Hierarchical Software Landscape Visualization for System Comprehension: A Controlled Experiment

Florian Fittkau, Alexander Krause, and Wilhelm Hasselbring  
Software Engineering Group, Kiel University, Kiel, Germany  
Email: {ffi, akr, wha}@informatik.uni-kiel.de

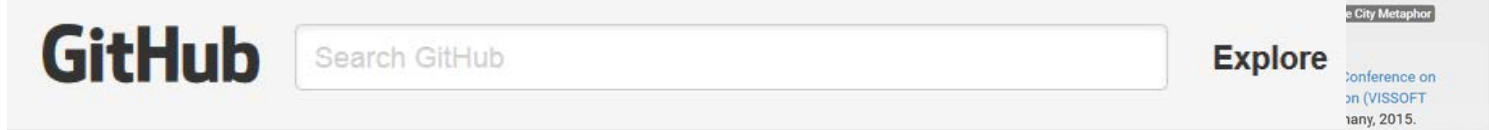
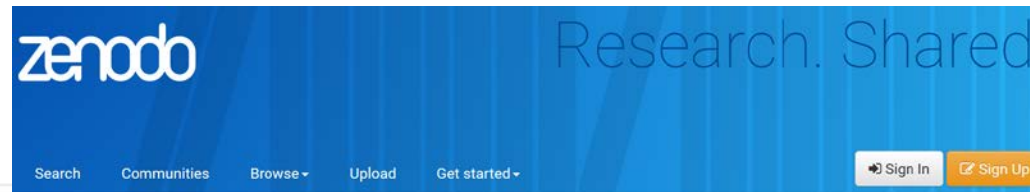


### Exploring Software Cities in Virtual Reality

Florian Fittkau, Alexander Krause, and Wilhelm Hasselbring

(Kiel University, Germany)

Preprint Available Video Info



## ExplorViz

Live trace visualization for large software landscapes

<http://www.explorviz.net>

22.03.2017

27

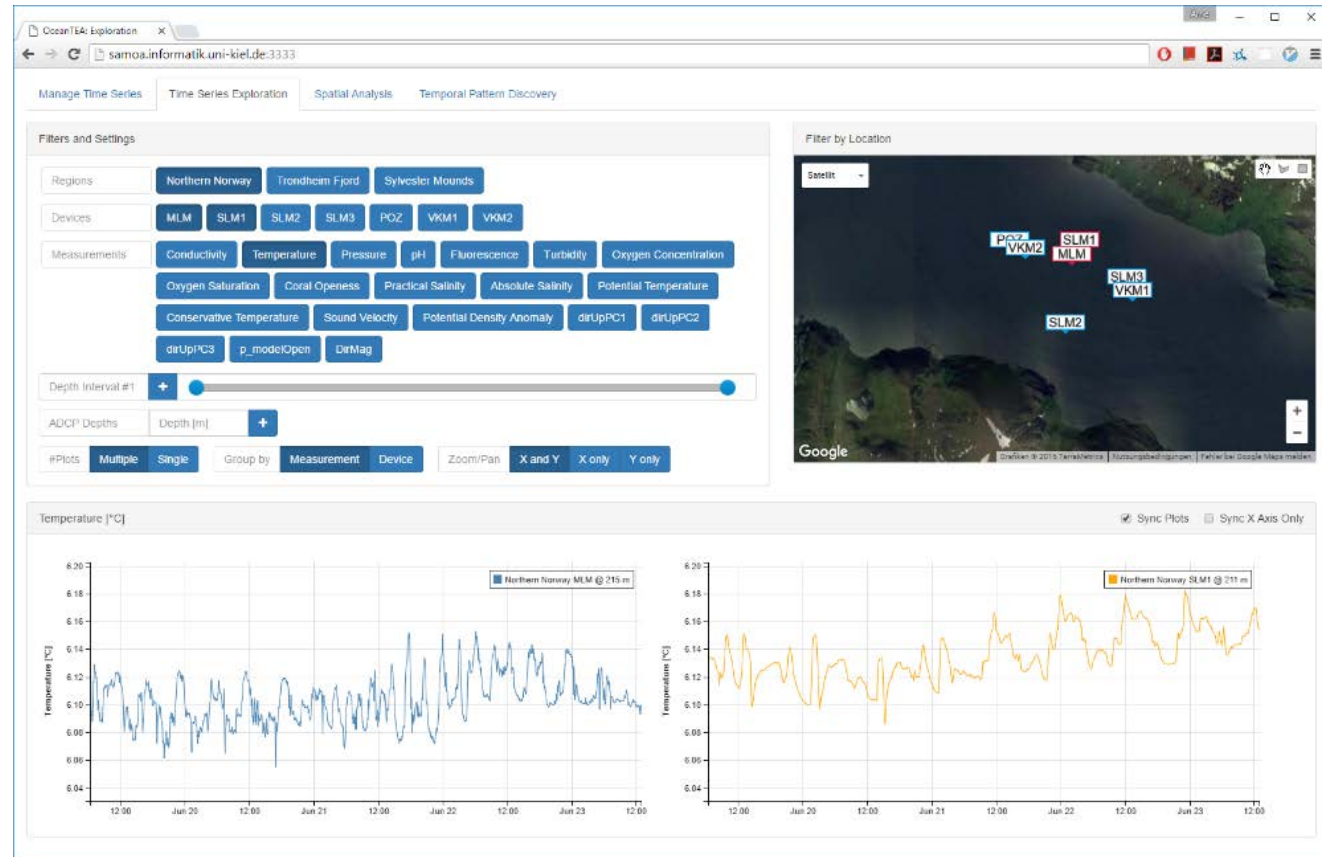
[Fittkau et al. 2013, 2015a-c, 2016]

# Cloud-Based Platform for Repeatable Ocean Observation Data Processing

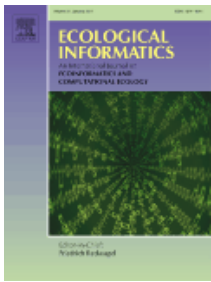
OceanTEA



future ocean  
KIEL MARINE SCIENCES



<https://github.com/a-johanson/oceantea> [Johanson et al. 2016a]



## Publishing:

- Paper: <https://www.journals.elsevier.com/ecological-informatics/>
- Code: <https://github.com/a-johanson/oceantea>
- Software service with data: <http://maui.se.informatik.uni-kiel.de:9090/>

### Modeling Polyp Activity of *Paragorgia arborea* Using Supervised Learning

Arne Johanson,<sup>a</sup> Sascha Flögel,<sup>b</sup> Wolf-Christian Dullo,<sup>b</sup>  
Peter Linke,<sup>b</sup> Wilhelm Hasselbring<sup>a</sup>

<sup>a</sup> Software Engineering Group, Kiel University, Germany  
<sup>b</sup> GEOMAR Helmholtz Centre for Ocean Research, Kiel, Germany

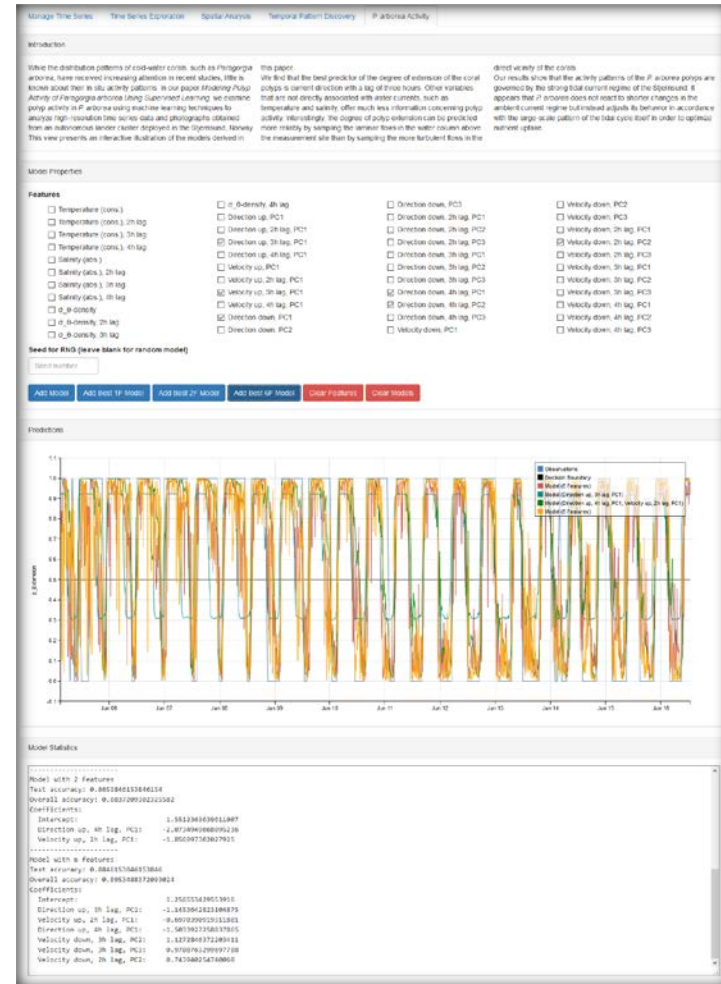
**Abstract**—While the distribution patterns of cold-water corals, such as *Paragorgia arborea*, have received increasing attention in recent studies, little is known about their *in situ* activity patterns. In this paper, we examine polyp activity in *P. arborea* using machine learning techniques to analyze high-resolution time series data and photographs obtained from an autonomous lander cluster deployed in the Stjernsund, Norway. An interactive illustration of the models derived in this paper is provided online as supplementary material.

We find that the best predictor of the degree of extension of the coral polyps is current direction with a lag of three hours. Other variables that are not directly associated with water currents, such as temperature and salinity, offer much less information concerning polyp activity. Interestingly, the degree of polyp extension can be predicted more reliably by sampling the laminar flows in the water column above the measurement site than by sampling the more turbulent flows in the direct vicinity of the corals.

Our results show that the activity patterns of the *P. arborea* polyps are governed by the strong tidal current regime of the Stjernsund. It appears that *P. arborea* does not react to shorter changes in the ambient current regime but instead adjusts its behavior in accordance with the large-scale pattern of the tidal cycle itself in order to optimize nutrient uptake.

#### 1 Introduction







Cold-water corals (CWCs) such as *Paragorgia arborea* and *Lophelia pertusa* can be found on continental shelves, slopes, and seamounts all over the world. Like tropical coral reefs, which inhabit shallower and warmer waters, CWC reefs are associated with high biodiversity as they provide habitat to many other species (Roberts and Cairns 2014; Roberts et al. 2006).





# Generic Research Data Infrastructure

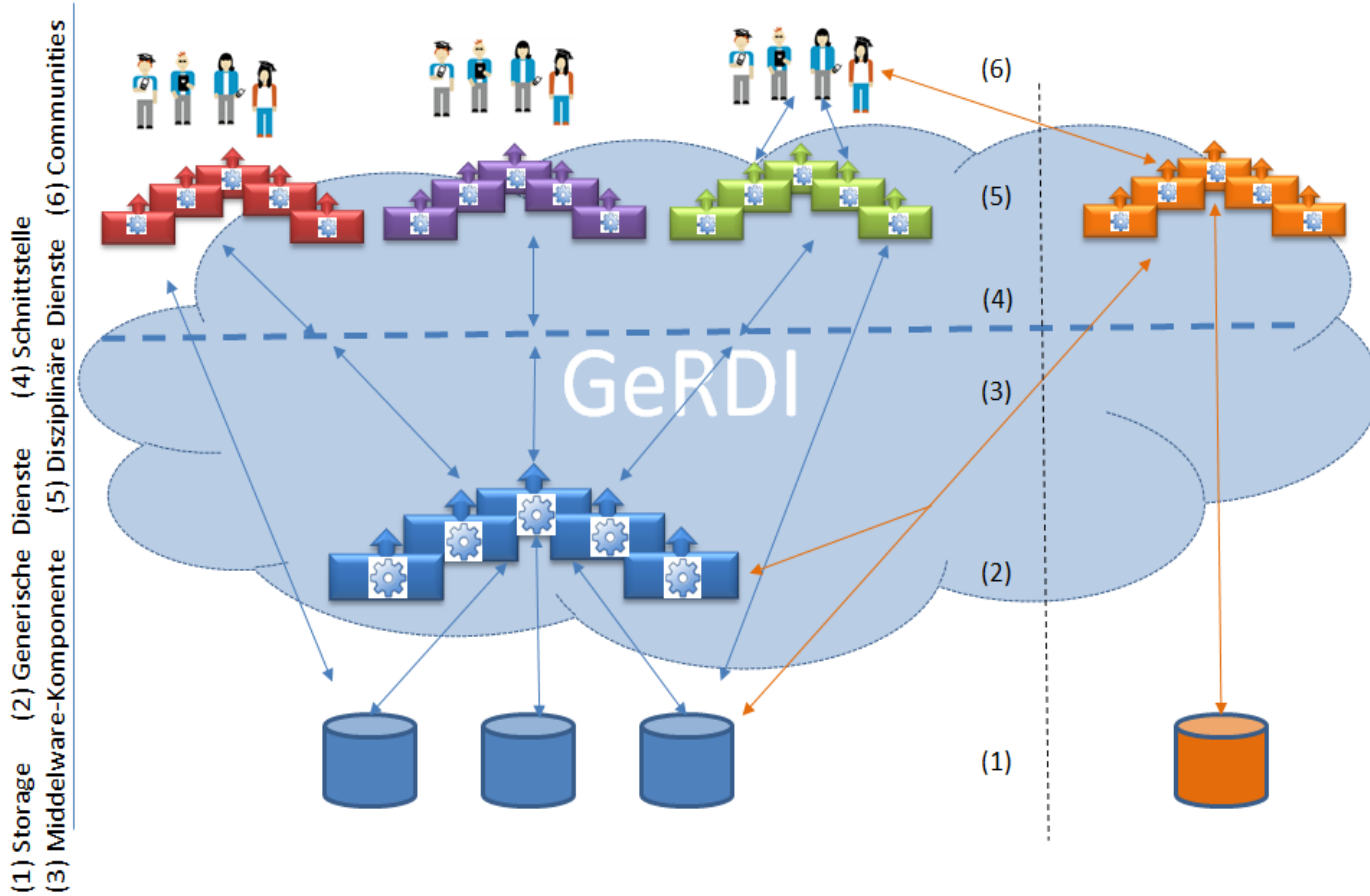


<p>Economics</p> 	<p>Life science, Humanities</p>  
<p>Marine science</p>  	<p>Environmental science</p> 



<http://www.gerdi-project.de/>

# Envisioned GeRDI Architecture



# 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>

# 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
- Outlook:
  - “Digital Ocean” in “Future Ocean Sustainability”

You may find these slides at: <http://eprints.uni-kiel.de/37072>

# References

- [Brauer & Hasselbring 2013] P. C. Brauer, W. Hasselbring: “PubFlow: a scientific data publication framework for marine science”. In: International Conference on Marine Data and Information Systems (IMDIS 2013), 2013, Lucca, Italy.
- [Fittkau et al. 2013] F. Fittkau, J. Waller, C. Wulf, W. Hasselbring: “Live Trace Visualization for Comprehending Large Software Landscapes: The ExplorViz Approach“, In: 1st IEEE International Working Conference on Software Visualization (VISOFT 2013).
- [Fittkau et al. 2015a] F. Fittkau, S. Roth, W. Hasselbring: “ExplorViz: Visual Runtime Behavior Analysis of Enterprise Application Landscapes“, In: 23rd European Conference on Information Systems (ECIS 2015).
- [Fittkau et al. 2015b] F. Fittkau, A. Krause, W. Hasselbring: “Hierarchical Software Landscape Visualization for System Comprehension: A Controlled Experiment”. In: 3rd IEEE Working Conference on Software Visualization, 2015.
- [Fittkau et al. 2015c] F. Fittkau, A. Krause, W. Hasselbring: “Exploring Software Cities in Virtual Reality“, In: 3rd IEEE Working Conference on Software Visualization, September 2015, Bremen, Germany.
- [Fittkau et al. 2016] F. Fittkau, A. Krause, W. Hasselbring: “Software landscape and application visualization for system comprehension with ExplorViz“, In: Information and Software Technology. DOI 10.1016/j.infsof.2016.07.004
- [Hasselbring 1994] W. Hasselbring: “Prototyping Parallel Algorithms in a Set-Oriented Language, Verlag Dr. Kovac, 1994.
- [Hasselbring 2000] W. Hasselbring: “Programming Languages and Systems for Prototyping Concurrent Applications“, In: ACM Computing Surveys, 31 (1). pp. 43-79. DOI 10.1145/349194.349199.
- [Hasselbring 2002] W. Hasselbring: “Component-Based Software Engineering“, In: Handbook of Software Engineering and Knowledge Engineering. World Scientific Publishing, Singapore, pp. 289-305, 2002.
- [Hasselbring 2006] W. Hasselbring: “Software-Architektur“. In: Informatik-Spektrum 29(1): 48-52, February 2006.
- [Hasselbring & Steinacker 2017] W. Hasselbring, G. Steinacker: “Microservice Architectures for Scalability, Agility and Reliability in E-Commerce“, In: Proceedings of the IEEE International Conference on Software Architecture (ICSA 2017), April 2017, Gothenburg, Sweden.
- [Heinrich et al. 2014] R. Heinrich, E. Schmieders, R. Jung, K. Rostami, A. Metzger, W. Hasselbring, R. Reussner, K. Pohl: “Integrating Run-Time Observations and Design Component Models for Cloud System Analysis“, In: 9th Workshop on Models@run.time 2014.

# References

- [Heinrich et al. 2015] R. Heinrich, R. Jung, E. Schmieders, A. Metzger, W. Hasselbring, R. Reussner, K. Pohl: “Architectural Run-Time Models for Operator-in-the-Loop Adaptation of Cloud Applications”, In: 9th IEEE Symposium on the Maintenance and Evolution of Service-Oriented Systems and Cloud-Based Environments (MESOCA 2015).
- [Johanson et al. 2016a] A. Johanson, S. Flögel, C. Dullo, W. Hasselbring: “OceanTEA: Exploring Ocean-Derived Climate Data Using Microservices”. In: Sixth International Workshop on Climate Informatics (CI 2016), September 2016, Boulder, Colorado.
- [Johanson et al. 2016b] A. Johanson, W. Hasselbring, A. Oschlies, B. Worm: “Evaluating Hierarchical Domain-Specific Languages for Computational Science: Applying the Sprat Approach to a Marine Ecosystem Model”. In: Software Engineering for Science. CRC Presspp. 175-200. DOI 10.1201/9781315368924-9.
- [Johanson et al. 2017a] A. Johanson, A. Oschlies, W. Hasselbring, A. Worm: “SPRAT: A spatially-explicit marine ecosystem model based on population balance equations”, In: Ecological Modelling, DOI 10.1016/j.ecolmodel.2017.01.020.
- [Johanson et al. 2017b] A. Johanson, S. Flögel, C. Dullo, P. Linke, W. Hasselbring: “Modeling Polyp Activity of Paragorgia arborea Using Supervised Learning”, In: Ecological Informatics, Elsevier, 2017.
- [Johanson & Hasselbring 2016] A. Johanson, W. Hasselbring: “Effectiveness and efficiency of a domain-specific language for high-performance marine ecosystem simulation: a controlled experiment”, In: Empirical Software Engineering. DOI 10.1007/s10664-016-9483-z.
- [Ploski et al. 2007] J. Ploski, W. Hasselbring, S. Schwierz, J. Rehwinkel: “Introducing Version Control to Database-Centric Applications in a Small Enterprise”, In: IEEE Software, 24 (1). pp. 38-44. DOI 10.1109/MS.2007.17.
- [Reussner & Hasselbring 2008] R. Reussner, W. Hasselbring: “Handbuch der Software-Architektur”. dpunkt.verlag, 2<sup>nd</sup> edition, 2008.
- [Waller et al. 2015] J. Waller, N. Ehmke, W. Hasselbring: “Including Performance Benchmarks into Continuous Integration to Enable DevOps“, In: ACM SIGSOFT Software Engineering Notes, 40(2).
- [Wulf et al. 2016] C. Wulf, C.C. Wiechmann, W. Hasselbring: “Increasing the Throughput of Pipe-and-Filter Architectures by Integrating the Task Farm Parallelization Pattern”, In: 18th International ACM SIGSOFT Symposium on Component-Based Software Engineering, April 2016, Venice, Italy.