

eSciDoc-Days  
October 2011

# A Path to filled Archives

or

‘Hey dude, how bumpy is it going to be?’

Dirk Fleischer  
dfleischer@ifm-geomar.de



# Complaining, complaining...

Data sharing: Empty archives



*'So we spent all this money,  
we spent all this time,  
we got the software up and running,  
and then we said,*

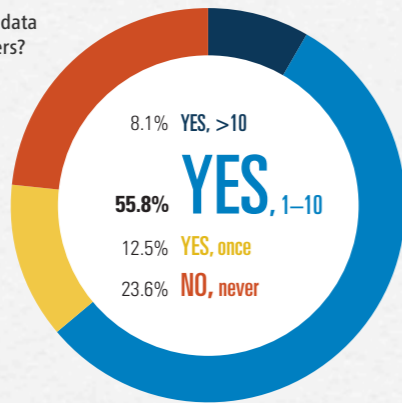
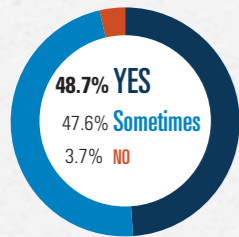
Bryn Nelson  
Nature 461, 160-163 (2009)  
<http://dx.doi.org/10.1038/461160a>

*'OK, here it is. We're ready.  
Give us your stuff" -  
"And that's where we hit the wall!"*

# Complaining, complaining...

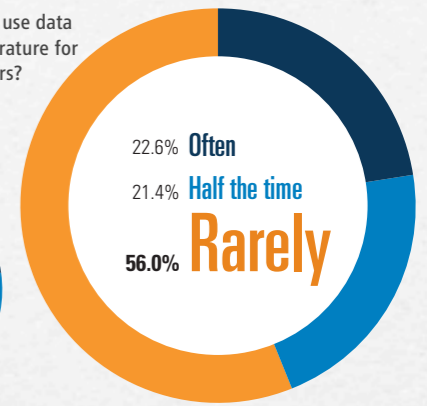
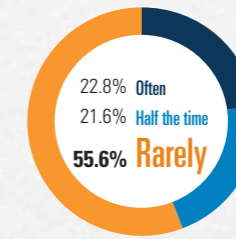
Have you asked colleagues for data related to their published papers?

If you answered yes, have the appropriate data been provided?



How often do you access or use data sets from the published literature for your original research papers?

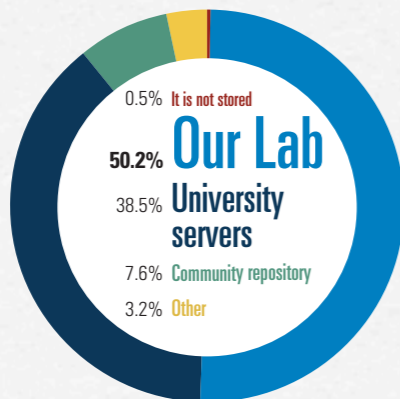
From archival databases?



Yael Fitzpatrick, using [www.wordle.net](http://www.wordle.net)  
Science, Vol. 331, Issue 6018

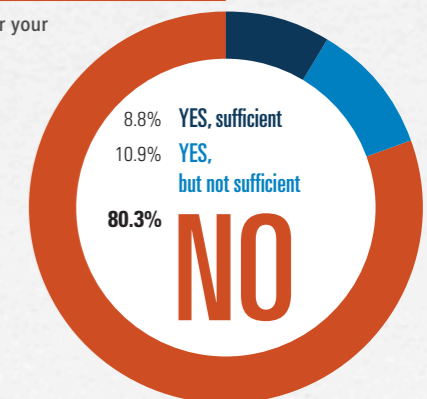
Where do you archive most of the data generated in your lab or for your research?

“Even within a single institution there are no standards for storing data, so each lab, or often each fellow, uses ad hoc approaches.”



Is there sufficient funding for your lab or research group for data curation?

“There are many tales of early archaeologists burning wood from the ruins to make coffee. If we fail to curate the environmental archives we collect from nature at public expense, we essentially repeat those mistakes.”



# Data Sharing?

Most researchers happily embrace the idea of sharing.

But in practice those advantages often fail to outweigh researchers' concerns.



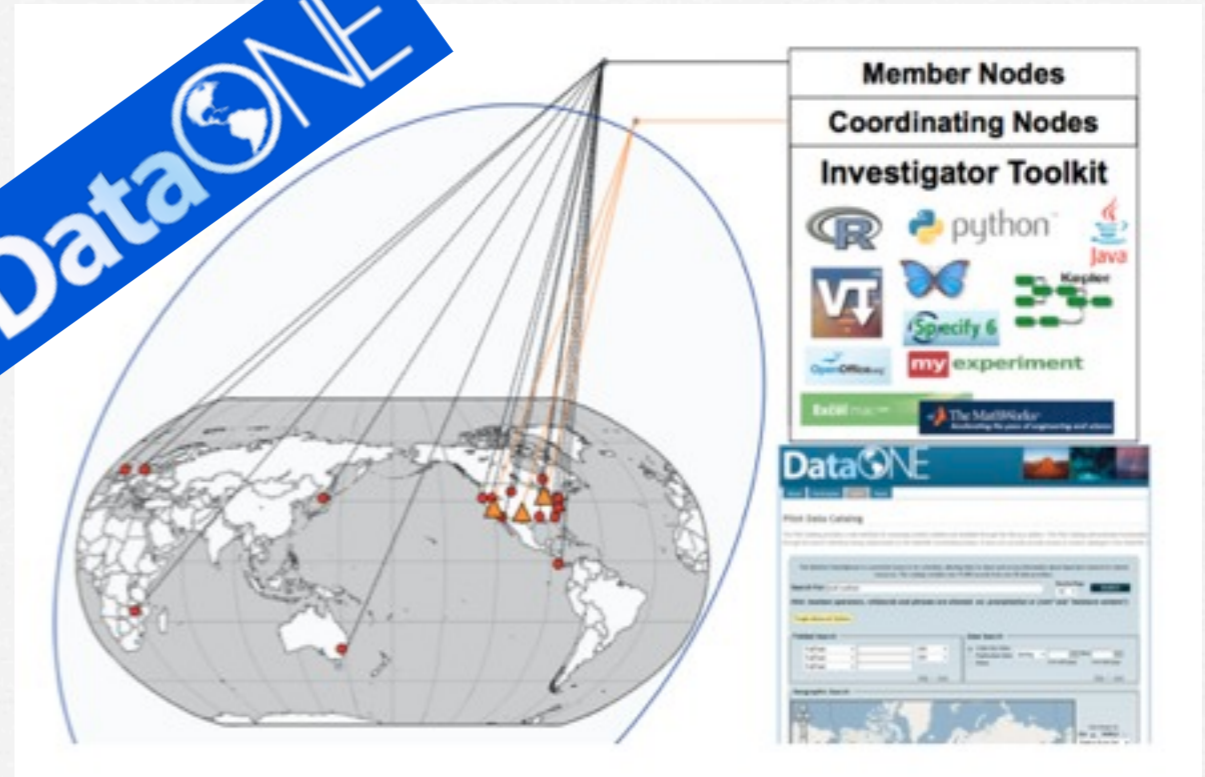
It opens up observations to independent scrutiny, fosters new collaborations and encourages further discoveries in old data sets.

"The phrase 'if you build it, they will come' does not yet apply to institutional repositories."  
A similar reality check has greeted other data-sharing efforts.

# The Carrot Crusade



Data Conservancy  
A Blueprint for Research Libraries



Helping you to find,  
access, and reuse data

W. Michener, 2011  
(D-Lib Magazine Vol. 17)

# The Carrot Crusade

Marine Micropaleontology  
Volume 66, Issues 3-4, 20 February 2008, Pages 192-207

Result list | previous < 1 of 1 > next  
Font Size:

Article | Figures/Tables | References | PDF (743 K)

Thumbnail | Full-Size Images

doi:10.1016/j.marmicro.2007.09.002  
Cite or Link Using DOI

Copyright © 2007 Elsevier B.V. All rights reserved.

## Organic matter rain rates, oxygen availability, and vital effects from benthic foraminiferal $\delta^{13}\text{C}$ in the historic Skagerrak, North Sea

Sylvia Gröbner<sup>a</sup> and Andreas Heekeren<sup>a</sup>

<sup>a</sup>Allied Wegener Institute for Polar and Marine Research, Columbusstr., D-27568 Bremerhaven, Germany

Received 27 March 2007; revised 21 September 2007; accepted 24 September 2007. Available online 4 October 2007.

### Abstract

The sediment cores 225514 and 225510 were recovered from 420 and 285 m water depth, respectively. They were investigated for their benthic foraminiferal  $\delta^{13}\text{C}$  during the last 500 years. Both cores were recovered from the southern flank of the Skagerrak. The  $\delta^{13}\text{C}$  values of *Uvigerina mediterranea* and other shallow infaunal species in both cores indicate that organic matter rain rates to the seafloor varied around a mean value until approximately AD 1950 after which they increased. This increase might result from changes in the North Atlantic Current System and a co-occurring persistently high North Atlantic Oscillation index state in the 1980s to 1990s, rather than from anthropogenic eutrophication. Using  $\delta^{13}\text{C}$  mean values of multiple species, we reconstruct  $\delta^{13}\text{C}$  gradients of dissolved inorganic carbon (DIC) within pore waters for the time periods AD 1500 to 1950 and AD 1950 to 2000. The calculated  $\delta^{13}\text{C}_{\text{DIC}}$  ranges, interpreted as indicating total organic matter remineralization due to respiration, are generally bigger in Core 225514 than in Core 225510. Since mean  $\delta^{13}\text{C}$  values of *U. mediterranea* suggest that organic matter rain rates were similar at

### Article Toolbox

- E-mail Article
- Export Citation
- Cited By
- Add to my Quick Links
- Save as Citation Alert
- Permissions & Reprints
- Citation Feed
- Cited By in Scopus (0)

### Related Articles in ScienceDirect

- Preface *Marine Geology*
- Living (stained) benthic foraminifera within surficial ... *Marine Geology*
- Dimorphic forms of recent foraminifera: An additional L. *Paleogeography, Paleoclimatology, Paleoecology*
- Holocene sedimentation in the Skagerrak: A review *Marine Geology*
- Mineralogy and geochemistry of a core from the Skagerra... *Marine Geology*

View More Related Articles

Supplementary Data

View Record in Scopus

settling  
ic matter: Source, diagenetic, or

1882 U.S.A.  
online 15 April 2003.



### Abstract

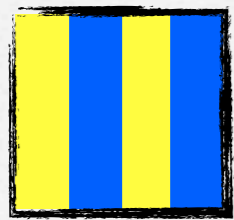
Geochemical studies of Cretaceous strata rich in organic carbon (OC) from Deep Sea Drilling Project (DSDP) sites and several land sections reveal several consistent relationships among amount of OC, hydrocarbon generating potential of kerogen (measured by pyrolysis as the hydrogen index, HI), and the isotopic composition of the OC. First, there is a positive correlation between HI and OC in strata that contain more than about 1% OC. Second, percent OC and HI often are negatively correlated with carbon isotopic composition ( $\delta^{13}\text{C}$ ) of kerogen. The relationship between HI and OC indicates that as the amount of organic matter increases, this organic matter tends to be more lipid rich reflecting the marine source of the organic matter. Cretaceous samples that contain predominantly marine organic matter tend to be isotopically lighter than those that contain predominantly terrestrial organic matter. Average  $\delta^{13}\text{C}$  values for organic matter from most Cretaceous sites are between  $-26$  and  $-28\text{‰}$ , and values heavier than about  $-25\text{‰}$  occur at very few sites. Most of the  $\delta^{13}\text{C}$  values of Miocene to Holocene OC-rich strata and modern marine plankton are between  $-16$  to  $-23\text{‰}$ . Values of  $\delta^{13}\text{C}$  of modern terrestrial organic matter are mostly between  $-23$  and  $-33\text{‰}$ . The depletion of terrestrial OC in  $^{13}\text{C}$  relative to marine planktonic OC is the basis for numerous statements in the literature that isotopically light Cretaceous organic matter is of terrestrial origin, even though other organic geochemical and/or optical indicators show that the organic matter is mainly of marine origin. A difference of about 5‰ in  $\delta^{13}\text{C}$  between modern and Cretaceous OC-rich marine strata suggests either that Cretaceous marine planktonic organic matter had the same isotopic signature as modern marine plankton and that signature has been changed by diagenesis, or

### Related Articles

- A synthesis of organic petrographic and geochemical res... *Physics and Chemistry of The Earth*
  - The Magnetic Epoch-6 carbon shift: A change in the ocea... *Marine Micropaleontology*
  - Erratum *Marine Micropaleontology*
  - $^{15}\text{N}/^{14}\text{N}$  variations in Cretaceous Atlantic sedimentary s... *Earth and Planetary Science Letters*
  - Polynofacies investigation of Cretaceous (Middle Jurassic... *Marine and Petroleum Geology*
- View more related articles

# The Carrot Crusade

- We will bring your data to the world, but before this there is something we would like you to do:



wrap them in blue and yellow paper

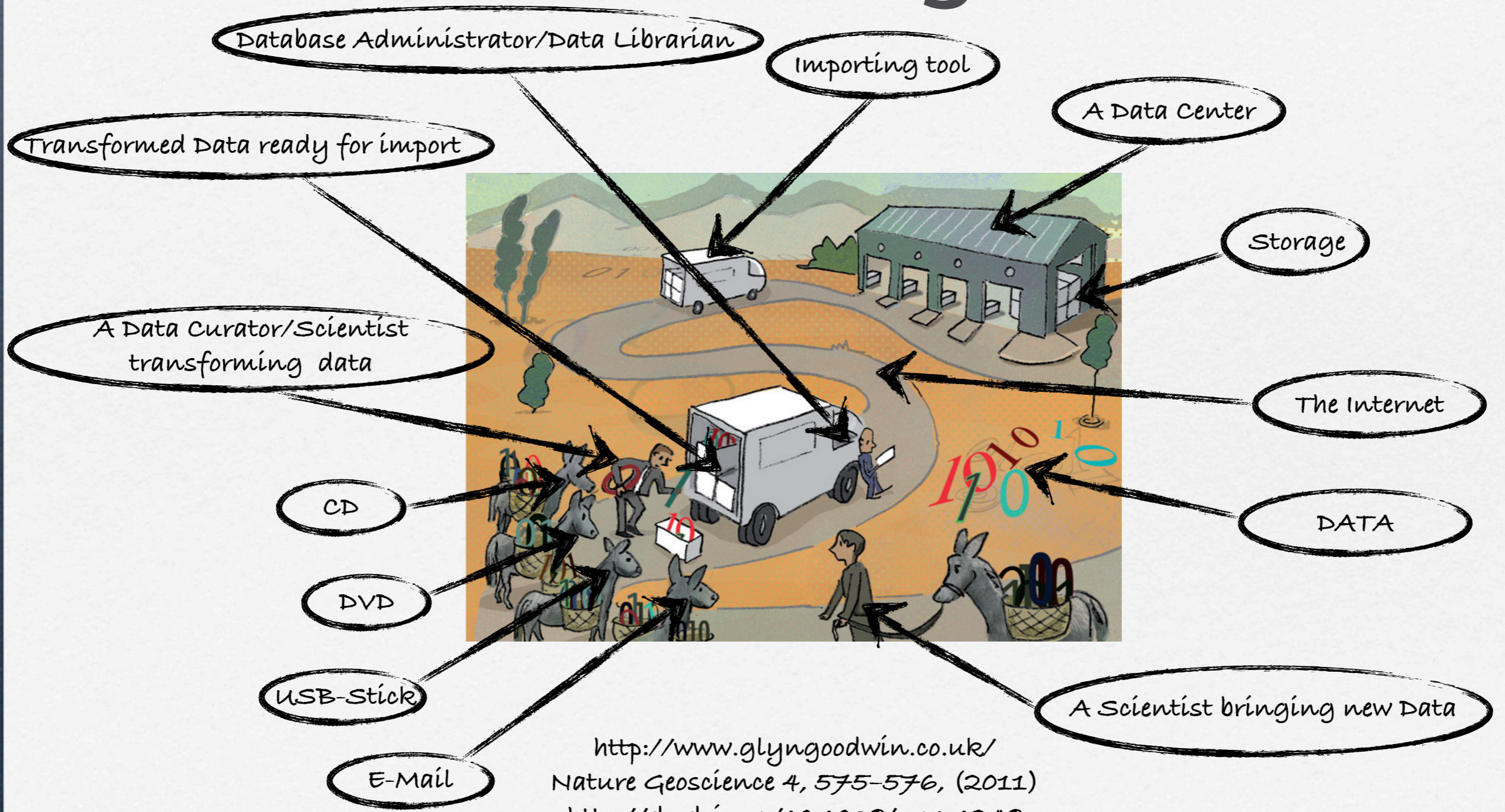


put green stickers on it best are stars, but flowers are also okay



don't forget the purple ribbon around it

# What's wrong here?



<http://www.glyngoodwin.co.uk/>  
Nature Geoscience 4, 575-576, (2011)  
<http://dx.doi.org/10.1038/ngeo1248>



# The Bottleneck!

- 700 Publications per year
- 3-4 days per import
- 260 working days per year

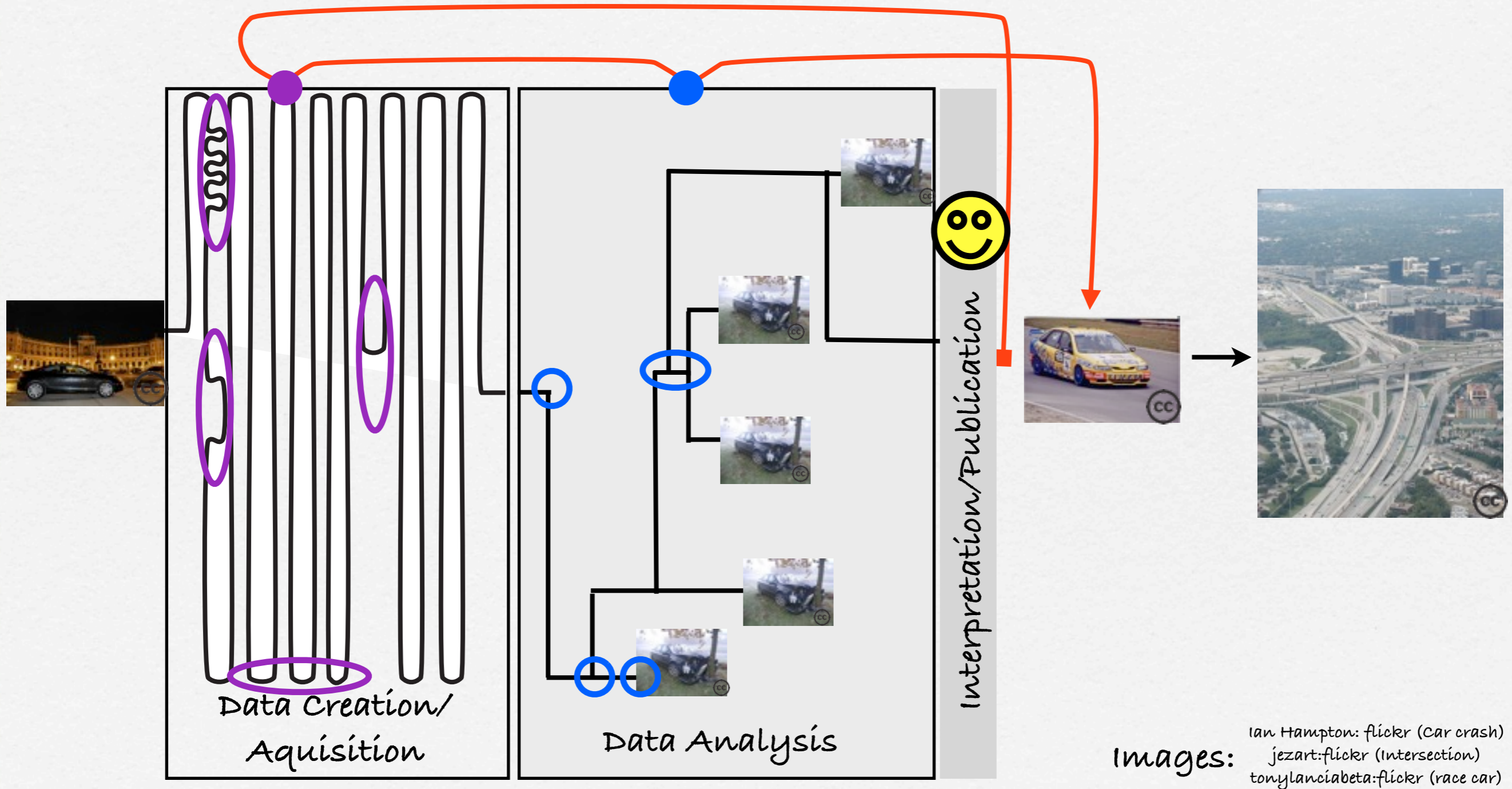
$$700 * 3 = \frac{2100}{260} = 8,07$$

$$700 * 4 = \frac{2800}{260} = 10,77$$

8 to 11 Data Managers

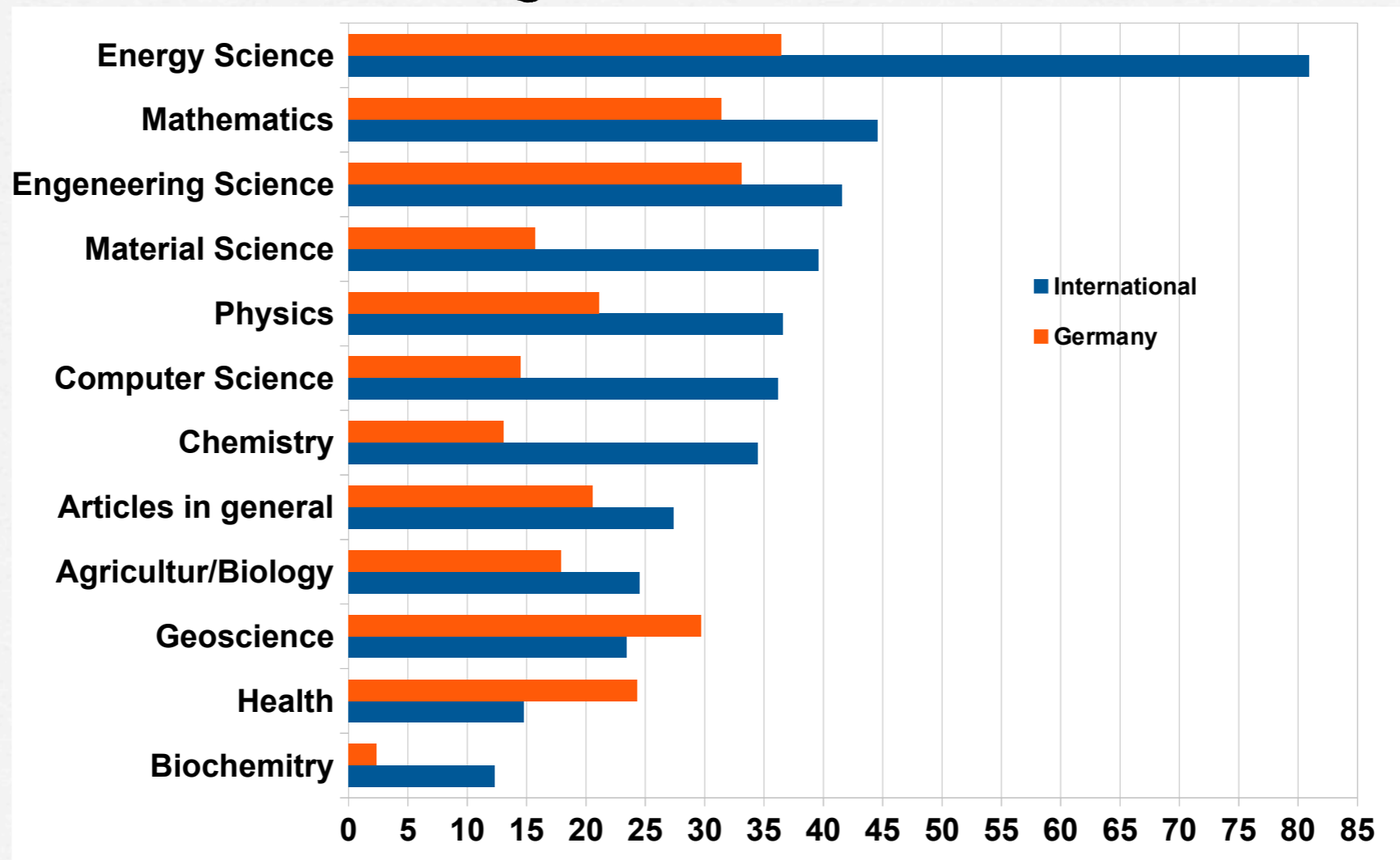
*If it could be done in TWO days  
You still need 5-6 Data Managers*

# Scientific Road Trip



# Publication Output

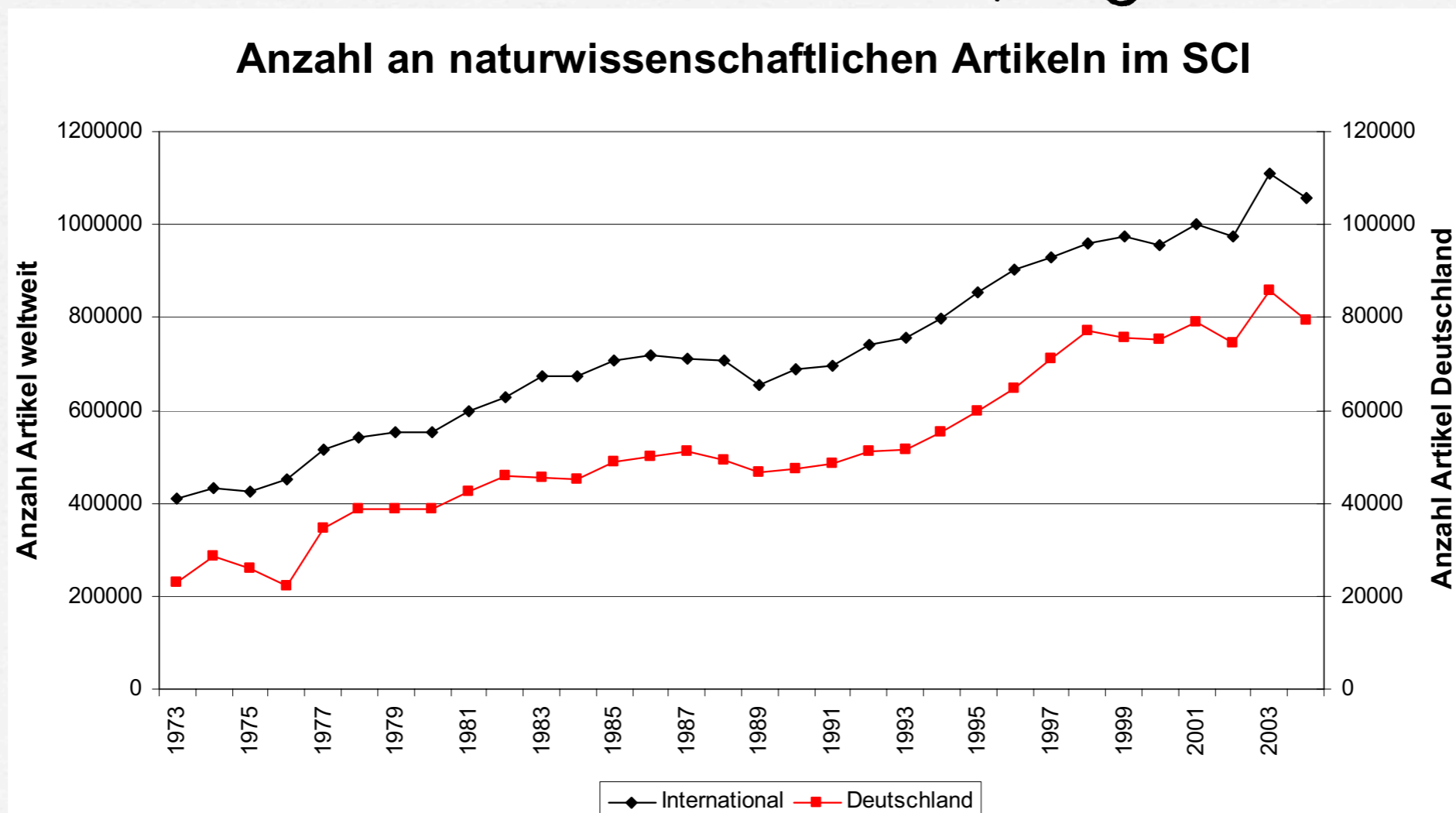
Relative growth of publication output from 1994-1999 to 2000-2004 by disciplines (SCOPUS)



Source: D. Tunger 2009 Forschungszentrum Jülich)

# Publication Output

Number of new scientific articles in  
Science Citation Index per year



Source: D. Tunger 2009 Forschungszentrum Jülich)

# What's next?

If this is all going to happen we definitely need technical support to reduce human interactions!

~~Institutions~~ Projects need to take action

Research  
Projects



Data Archive



~~Institutions~~

# Projects need to take action

Data Provenance

Research Site

Information

Research Site

Research Site

Research Site

Research Site

Data Archive

Retrievability

usability

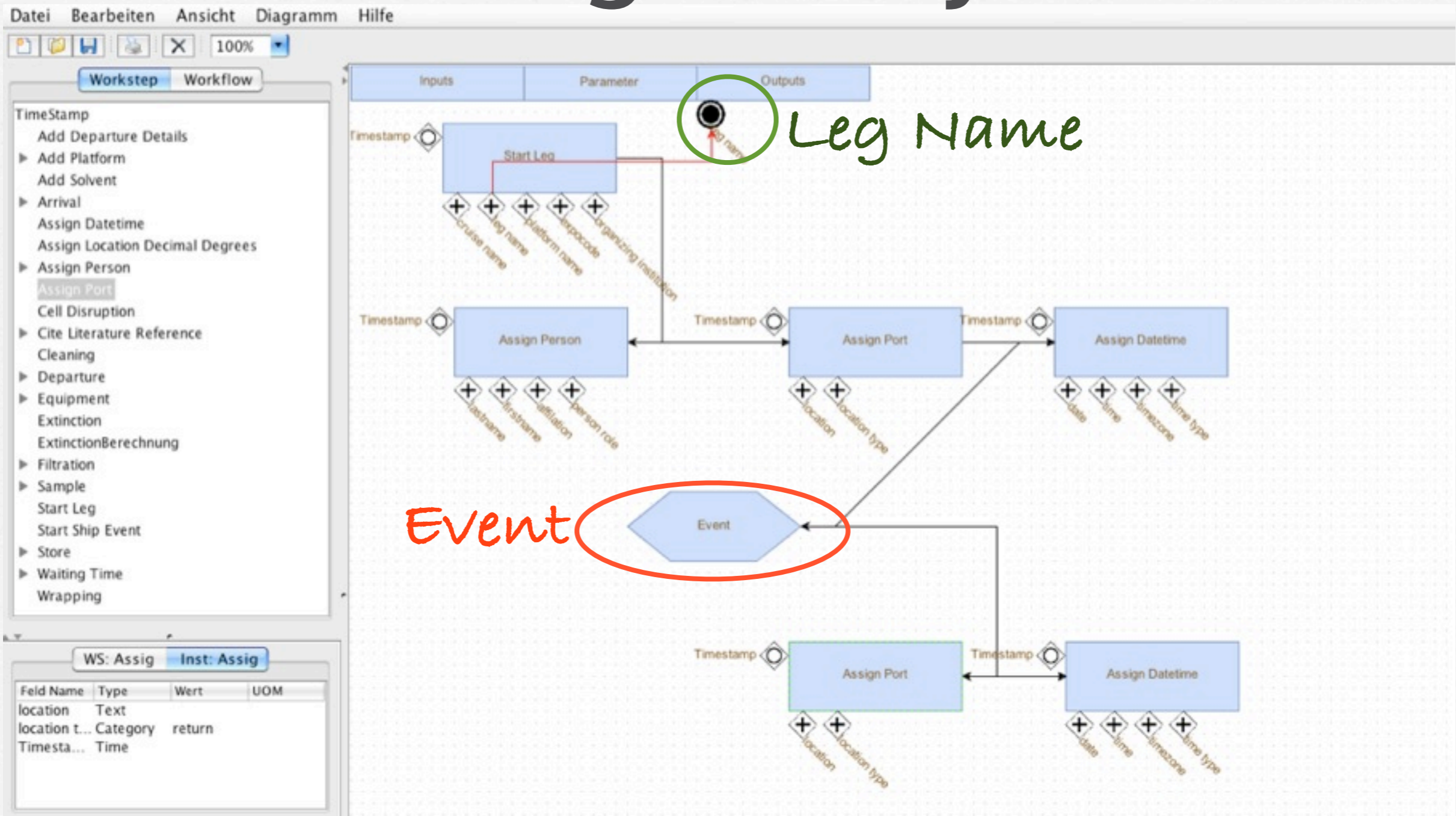


# Why Research Sites?

- Personal and short communication between Scientist and Data Center staff
- Sustainability of trusted personal cooperation
- Scientific record of the performed research history for a site (University, Institutes, etc.)
- Data capturing at the point of origin
  - Collecting the unpublished data
  - Capturing data and meta information at the time of data creation
  - Storing the analytic procedures as provenance information
  - Publish data with on-click solution from structured data source to another



# Modeling what you do



# Human activities

The screenshot shows a web browser window with the URL <https://portal.ifm-geomar.de/group/data-management/datenman>. The browser's address bar and search bar are visible. Below the browser window, a workflow diagram is overlaid on the page content. The diagram consists of several colored boxes and buttons representing different activities in a process. The main content area of the browser shows a form for 'Event 2' with fields for 'Start Ship Event', 'Leg Name' (M77/3), 'Description', 'Station Number', and 'Gear Name'. A red arrow points to the 'Assign Person' button in the workflow diagram. To the right of the 'Event 2' section is a 'Repetitions' grid with buttons numbered 1 through 14, and an ellipsis button.

**Start Leg**

**Assign Person**

**Assign Port**

**Assign Date/Time**

**Event 2**

**Start Ship Event** Leg Name: M77/3

Description:

Station Number:

Gear Name:

**Assign Person**

**Assign Date/Time**

**Assign Location Decimal Deg.**

**Repetitions**

1 2 3

4 5 6

7 8 9

10 11 12

13 14 ...

**Assign Port**

**Assign Date/Time**

<https://portal.ifm-geomar.de/group/data-management/calendar>



ozean der zukunft  
DIE KIELER MEERESWISSENSCHAFTEN

# Thank you!



IFM-GEOMAR

- Thanks to:
  - Andreas, Carsten, Hela and Pina
  - Kai Jannaschk, B. Thalheim
  - Computer Science Department Kiel
  - Funding Projects



SFB 754

Dirk Fleischer  
dfleischer@ifm-geomar.de

