

# Open Access and Open Data in Germany: current political developments vs. grassroots approaches

Anita Eppelin, German National Library of Medicine

Open Access – Open Data Conference

Cologne, 13th Dec 2010



## Before we start ...

- Fourth paradigm
- Discipline (and other) specifics
- Incentives, awareness
- Responsibilities, infrastructure, support
- Funding, business models
- Copyright





# Agenda

- **General framework for science**
- Network of players on different levels; their interests, strategies and approaches
- Two oppositional approaches in detail:
  - Official level: „Commission for the future of the information infrastructure”
  - Grassroots level
- Conclusion



## Framework for science: general

- „Freedom of science“, principle of self-organization
- Public budgets' need to economise
- Strong copyright law
- Heterogeneity
- Paradigm shift toward digital scientific communication, data-driven science



## Framework for science: significant official positions

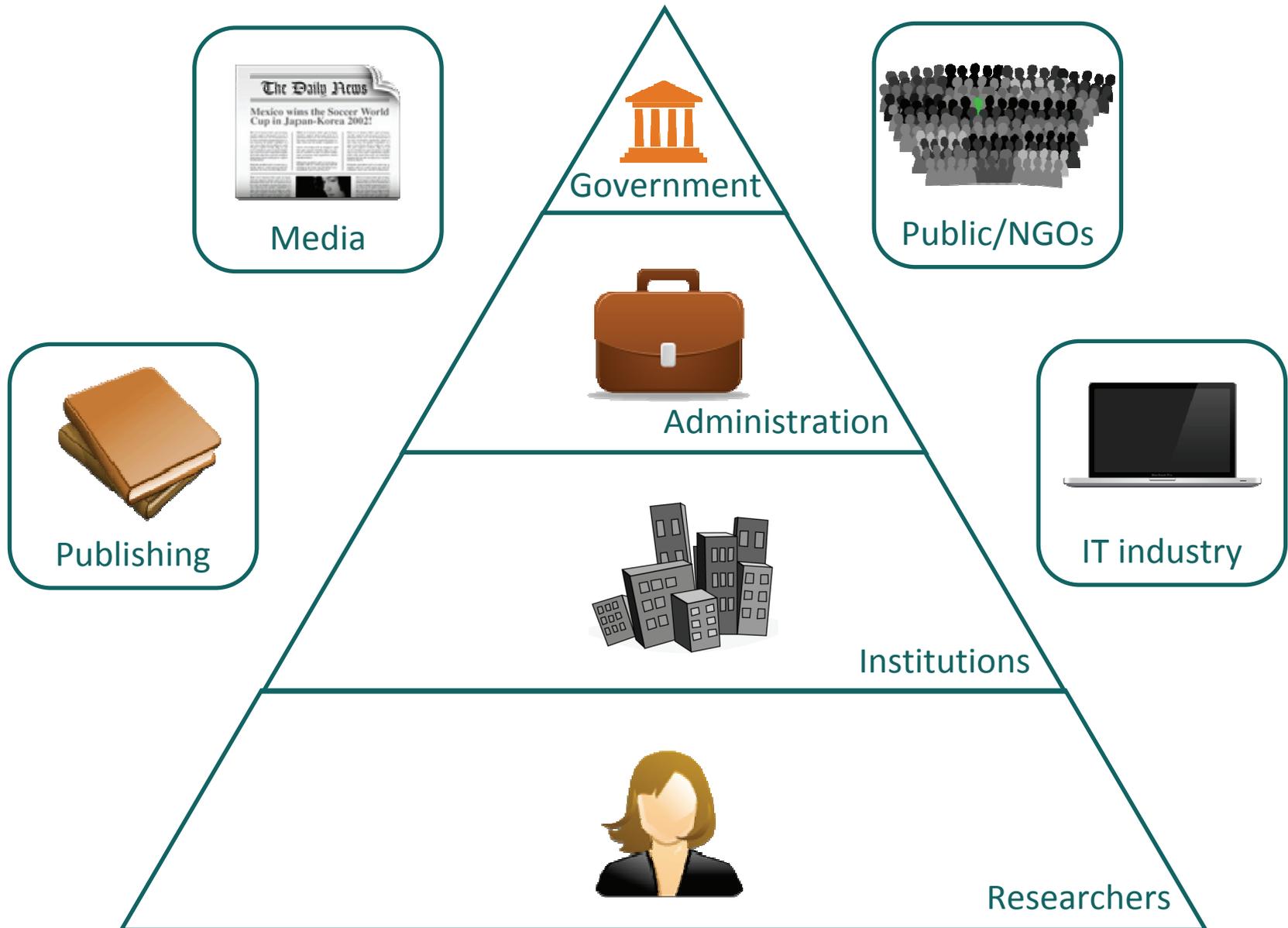
- BBB: Budapest, Bethesda, Berlin
- Göttingen Declaration on Copyright for Education and Research (2004)
- Funding bodies (DFG, ESF): funding programs and criteria regarding publications and data
- (Re)definition process: „Commission for the Future of the Information Infrastructure“ (2010)



# Agenda

- General framework for science
- **Network of players on different levels; their interests, strategies and approaches**
- Two oppositional approaches in detail:
  - Official level: „Commission for the future of the information infrastructure”
  - Grassroots level
- Conclusion

# Levels and players





# Government

- Strategy:
  - Balancing interests: support for both science and business location Germany
  - Funding of research in terms of self organization
- Approaches:
  - Revision of Copyright law (3rd basket)
  - Redefinition process underway („Commission for the future of the information infrastructure“)



## Science administration: „Alliance of German Research Organizations“

- Strategy: co-operation and networking
  - Coordination of activities of member organizations
  - Establishing guidelines, e.g. common „Data Policy“ (2010)
- Approach: Lifting activities to the next level, e.g. with comprehensive publications, workshops, surveys



## Science administration: funding bodies

- Strategy: targeted funding and active participation in developments through funding programs and underlying criteria
- Approaches:
  - Funding of OA green and gold projects, OA information platform on the web
  - Funding of Virtual Research Environment projects
  - Distinctive data funding strategy still under development



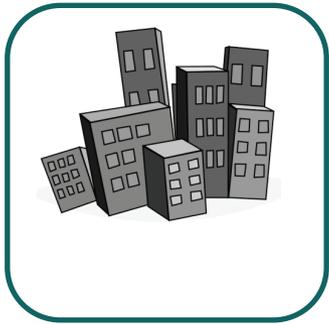
## Science administration: research organizations

- Strategy: coordination of activities of their member institutions
- Approaches:
  - Advocacy, information policy, consulting, e.g. Helmholtz policy and work program
  - Centralized or distributed infrastructures for OA/D:
    - Support, funding of publishing activities in general, also OA
    - Setup, establishing of subject-specific or institutional OA and data repositories



# Institutions: universities, research institutes

- Strategies vary, due to:
  - Autonomy vs. centralized organization, federalism
  - Policies of their umbrellas (impact?)
  - Infrastructure/support by libraries/IT to different extent
- Approaches:
  - Engaged in OA green, gold, VRE projects (mostly institutional approaches)
  - Disciplinary approaches by cooperations between universities and research institutions outside universities
  - Coalition for Action “Copyright for Education and Research”



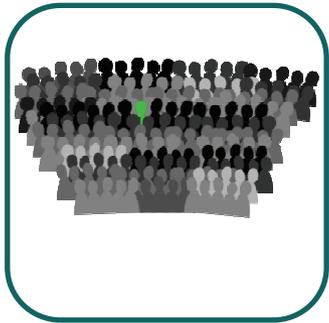
## Institutions: information infrastructure

- Strategies:
  - Redefinition of tasks, keeping up with developments
  - (Applied) research
- Approaches:
  - Setup, establishing, integration of infrastructures for whole research process: OA green & gold, data, long-term archiving
  - Education



# Individual researchers

- Strategies: interested in impact, career, leading figures within communities, minimizing effort
- National boundaries irrelevant
- Awareness for legal, financial & technical barriers rises
- Approaches:
  - „Petition for free purchase of scientific publications“
  - But also: oppositional initiatives like „Heidelberger Appell“
  - Engagement in OA/D initiatives as editors
  - Crowdsourcing: OpenWetWare, Dataverse, ResearchGate, ...



## The public, NGOs

- Nothing of similar impact like the US' taxpayer alliance or student's initiative
- Crowdsourced „Open Data“: mostly non-scientific (governmental or geospatial data)
- Open Data Network/Open Knowledge Foundation
- Focus on:
  - Data journalism
  - Creating services using existing data



# Media

- Strategy: often conservative position
  - Support for „Heidelberger Appell“, Swedish Chemists‘ initiative
  - General skepticism about digital scientific environment
  - Approaching decision-making units, triggering public debate
- Effects e.g. on copyright developments
- Awareness within internet/blogging community focusing on internet policy



## Publishing industry

- Initially: ignorance or even rejection of OA and data, due to lack of proper business models
- Meanwhile: acknowledged as profitable and sustainable sources of income
- OA and data publishing services are offered, conditions vary
- Still: lobbying for a Copyright law protecting publishers' interests



## IT industry

- Big market potential
- Building infrastructures for technical backbones (IBM, Dell, HP etc.)
- Developing and hosting platforms for data: archiving, sharing, analysis, visualization (Google, Microsoft, Amazon, Apple etc.)



# Agenda

- General framework for science
- Network of players on different levels; their interests, strategies and approaches
- **Two oppositional approaches in detail:**
  - **Official level: „Commission for the future of the information infrastructure”**
  - **Grassroots level**
- Conclusion



## „Commission for the future of information infrastructure“

- Government-initiated approach to find a common conceptual basis for future developments
- Not externally written, but by relevant involved parties
- Therefore neither genuinely top-down nor self-organized
- Aim: „a master plan that:
  - Depicts an optimized landscape of information infrastructure in Germany,
  - Describes the necessary framework conditions,
  - Points out possible synergy effects, cooperations, and release of resources.“



# „Commission for the future of the information infrastructure“

## Timeline:

- Spring 2009: Government commissions a concept for the future of the information infrastructure in Germany, Task Force is compiled
- Fall 2009: Task force suggests conceptual framework, government accepts
- Winter 2009: statutory meeting; members of Steering Committee: Alliance et al.
- Since February 2010: Working Groups prepare concept
- Spring/summer 2011: reaction from government



# „Commission for the future of the information infrastructure“

- Topics / Working Groups:
  - Licensing
  - Hosting/long-term archiving,
  - Non-textual materials
  - Retrodigitisation / cultural heritage
  - Virtual research environments
  - Open Access
  - Research data
  - Information literacy / higher education



# „Commission for the future of the information infrastructure“

Recommendations regarding research data and Open Access :

- Research data:
  - „should be permanently stored and provided to the (expert) public and future (researcher) generations for subsequent use in terms of a sovereign task.“
  - Setup and establishing of a national alliance, internationally linked
  - Establishing of reputation-relevant set of incentives for researchers
- Open Access:
  - „Institutions of scientific infrastructure are to build sustainable Open Access infrastructures based on disciplinary demands and integrate them in existing and emerging knowledge contexts.“



## Grassroots approaches

- Generally: impulse to act derives out of discipline-specific requirements of data-intensive research
- Therefore approaches often scientists-initiated
- When project establishes, infrastructure and administration tasks emerge, institutions and funding bodies get involved
- Infrastructure-driven approaches (might) have a very tough start or even fail



# Grassroots approaches: prototypical examples

- Open Access:
  - arXiv: prototypical grassroots initiative, institutionally sustained
  - ResearchGate: low threshold for contribution to Open Access pool, incentive: profiling, reputation management
- Data:
  - Pangea: priming during International Geophysical Year 1957 → WDC approach born by scientists
  - ESSD: advanced form of integration of data and publications



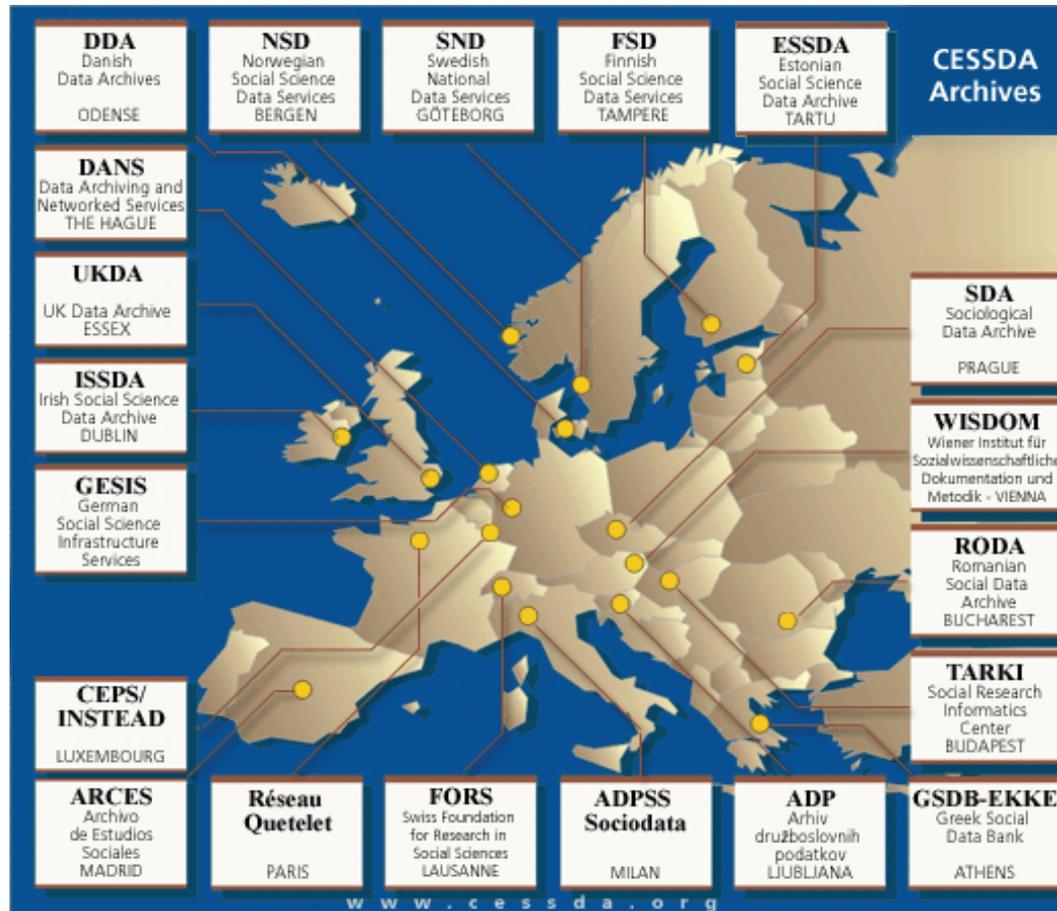
# Oceanography



**PANGAEA<sup>®</sup>**

**Publishing Network for Geoscientific & Environmental Data**

# Social sciences



→ Network of Social Science Data Archives

# Humanities

The screenshot shows the homepage of arts-humanities.net. At the top, there is a dark green header with the site logo on the left and a search bar on the right. Below the search bar are navigation links: About, Contact, Login/register, Projects, Methods, Tools, Library, Centres, Community, and News. The main content area features a section for the DH2010 conference, a sidebar with 'News and Feeds' and a list of categories like 'Browse & add events', and a central section titled 'arts-humanities.net: guide to digital humanities & arts' which includes a list of services and a 'Recently published' list of articles.

**arts-humanities.net**

Search

About | Contact | Login / register

Projects | Methods | Tools | Library | Centres | Community | News

**DH2010: Digital Humanities conference**

From 7th to 10th July, [DH2010](#), the annual international conference for digital scholarship in the humanities, was hosted by King's College London. arts-humanities.net functions as the *virtual conference*, giving you access to [video recordings](#) of lectures and events, [delegate interviews](#), an [image gallery](#) and a [conference blog](#).

**News and Feeds**

- Browse & add events
- Browse & add jobs
- Browse & add news
- Browse & add projects
- Browse & add tools
- Discussion forums
- Bibliography

**arts-humanities.net: guide to digital humanities & arts**

arts-humanities.net aims to support and advance the use and understanding of digital tools and methods for research and teaching in the arts and humanities by providing:

- Information on projects creating and using digital content, tools and methods to answer research questions
- Information on tools and methods for creating and using digital resources
- A listing of expert centres and individual researchers
- A library documenting lessons learned through case studies, briefing papers, and a bibliography

We invite you to [join as a member](#), and to contribute information about your projects, tools and research, publicise events and job vacancies, and take part in discussions. [\[read more\]](#)

[Browse by topic](#) [Browse by discipline](#)

**Recently published**

- Arts, Humanities, and Complex Networks – 2nd Leonardo satellite symposium at NetSci2011  
09/12/2010 - 00:51
- Research Associate (Digital Palaeography)  
08/12/2010 - 16:32
- Teaching Fellow in GIS  
07/12/2010 - 16:20
- Research databases in the humanities – where next?  
03/12/2010 - 16:34
- Corpus of Modern Scottish Writing (CMSW) launch  
30/11/2010 - 13:51
- CfP: OSGIS 2011  
30/11/2010 - 12:15

[more](#)

JISC Arts & Humanities Research Council KING'S COLLEGE LONDON

→ Digital Humanities community

# Epidemiology

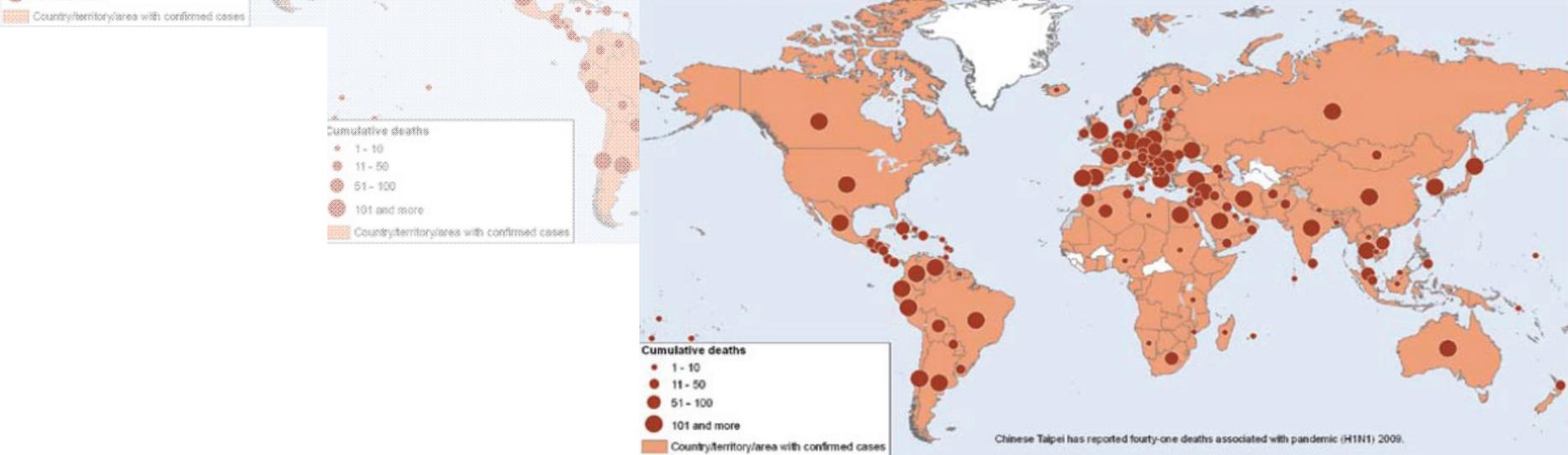
June 2009



Sep 2009



Mar 2010



## PLoS Currents: Influenza

A moderated collection for rapid and open sharing of useful new scientific data, analyses, and ideas.

The continuing public health burden due to influenza highlights the need for the rapid exchange of scientific results and ideas. PLoS Currents: Influenza aims to enable this exchange by providing an open-access online resource for immediate, open communication and discussion of new scientific data, analyses, and ideas in the field of influenza. All content is reviewed by a group of expert researchers.

[Link](#) [Citation](#) [Email](#) [Print](#) [Favorite](#) [Collect this page](#)

[Aims and Scope](#) | [Board of Reviewers](#) | [Author Guide](#) | [Reviewer Guide](#) | [FAQs](#) | [Creating a Google Account](#)

Sort by:   Reverse sort

- 

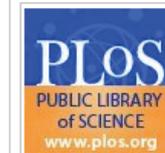
[Projection of seasonal influenza severity from sequence and serological data](#) Views: 1000  
Accepted: 4 days ago  
By [Yuri I Wolf](#), [Anastasia Nikolskaya](#), [Joshua L. Cherry](#), [Cecile Viboud](#) et al (6 authors)  
Severity of seasonal influenza A epidemics is related to the antigenic novelty of the predominant viral strains circulatin...
- 

[Tempo and Mode in the Molecular Evolution of Influenza C](#) Views: 100  
Accepted: Dec 1, 2010 9:20 AM  
By [Derek Gatherer](#)  
Abstract: Influenza C contributes to economic damage caused by working days lost through absence or inefficiency and may...
- 

[Fatal cases associated with pandemic influenza A \(H1N1\) reported in Greece.](#) Views: 700  
Accepted: Nov 10, 2010 8:44 AM  
By [Maria Athanasiou](#), [Theodore Lytras](#), [Georgia Spala](#), [Eleni Triantafyllou](#) et al (12 authors)  
ABSTRACT Between 18 May 2009 and 3 May 2010, a total of 149 fatal cases associated with pandemic influenza A (H1N1) were...
- 

[The 2009 pandemic H1N1 influenza vaccination in France: who accepted to](#) Views: 200  
Accepted: Oct 20, 2010 8:32 AM

[Edit this collection](#) [Write a knol](#)



**PLoS**  
The Public Library of Science (PLoS) is a non-profit open-access publisher, committed to making scientific and medical research literature a freely accessible public resource.

Article rating: ★★★★★  
Your rating:

[Moderator-centric collection](#)

[Moderated collaboration](#)

[Creative Commons Attribution 3.0 License](#)

[Top Viewed Knol Award](#)

[Top Pick Knol Award](#)

Version: 213

Last edited: 4 days ago.

Versions

Search for uses of this page ▼

### Categories

flu, h1n1, Health, Influenza

Based on community consensus.

[Sign in to add or vote for categories](#)

Learn more about categories

### Activity for this collection

This week:

1213 pageviews



# Herding cats: the sociology of data integration

Robert W. Williams<sup>1</sup>\*

<sup>1</sup> Department of Anatomy and Neurobiology, Center for Integrative and Translational Genomics, University of Tennessee Health Science Center, Memphis, TN, USA

\* Correspondence: [rwilliams@utmem.edu](mailto:rwilliams@utmem.edu)

*Upon this gifted age, in its dark hour,  
Rains from the sky a meteoric shower  
Of facts...they lie unquestioned, uncombined,  
Wisdom enough to leech of our ill*

*Is daily spun; but there exists no loom  
To weave it into fabric;*

—Edna St. Vincent Millay (from *Huntsman, What Quarry?* 1939)

## THE CHALLENGE OF PREDICTION

The age of personalized medicine and genomics is upon us and we are facing a grand challenge—or a brick wall. We have finally gained a near complete compendium of fundamental mechanisms, connections, and developmental processes: an encyclopedia of biology, bodies, brains, and behavior. We have the data, the density and integration needed to develop holistic and robust models that generate useful predictions? Will we be able to distinguish between personalized genomic ascos? What new types of resources, datasets, and analytic frameworks are needed to make correct diagnoses and recommend actions? What is my personal risk for Alzheimer's disease? How can I reduce my risk of it today?

The complexity of biological systems implies that a parts list of mechanisms and processes, however complete, will not be sufficient for the task of making good predictions. We have a very complex system, our models using a system that has the same level of complexity as human populations. I will describe an effective approach that

relies on genetic reference panels (GRPs) that can be used to make and test predictions from base pair to behavior. I will describe how scientists can retain their independence while explicitly contributing to a fabric of tightly woven quantitative data.

## THE COLLECTIVE COST OF SCIENTIFIC INDEPENDENCE

Scientists are trained to think independently and critically. It is inevitable that we like to do things our own way, generating and using data from experiments we designed ourselves. This approach is not a self-indulgent luxury—it is an essential attribute of innovative science, enshrined in the ways we evaluate and fund new and ongoing research. Independence contributes to the stirring cacophony of competing ideas that motivates our deepest understanding of biological processes.

Yet independence has a cost. The scope of studies from single genes to entire genomes, from single individuals to entire populations, and from modest budgets to the collective result is a fragmented, half-hidden literature and a fragmentary and rapidly evaporating collection of data. We have a vast amount of data, but it is scattered and under different conditions, tested using varied paradigms, and measured using different equipment. Of course the pieces do not fit together. It is difficult to even begin to integrate the data. The challenge of integrating the data is not just that the data are scattered and measured under different conditions, but that the data are often generated under different conditions, tested using varied paradigms, and measured using different equipment. Ronald Fisher pointed out that “the cost of a single experiment is often a very small fraction of the cost of a series of experiments” (Fisher, 1935). Fisher pointed this out in the context of scientific

“It is relatively easy to build a raft to float across a wide river. [...] To build a permanent bridge [it] requires the output of industries, the work of many specialists. [...] It is time to begin building the massive data bridges. [...] Perhaps the real challenge will be convincing both the community and its leaders that it is doable today.”





# Agenda

- General framework for science
- Network of players on different levels; their interests, strategies and approaches
- Two oppositional approaches in detail:
  - Official level: „Commission for the future of the information infrastructure”
  - Grassroots level
- **Conclusion**



## Conclusions

- Catalyzing between grassroots and strategic level could speed up establishing and integration of infrastructures.
- Disciplinary approaches are able to reach critical mass.
- There is little interconnectivity between different disciplinary approaches beyond communities.
- Do OA and OD efforts have the potential to promote each other?

Thank you for your attention!

Anita Eppelin  
eppelin@zbmed.de