



# Web networks of collaboration

Isidro F. Aguillo

Cybermetrics Lab. CCHS – CSIC

[isidro.aguillo@cchs.csic.es](mailto:isidro.aguillo@cchs.csic.es)



EXPLORING THE EMERGENT EUROPEAN RESEARCH SYSTEM

SEVILLE, 16-17 NOVEMBER 2009

# ERA Vision 2020

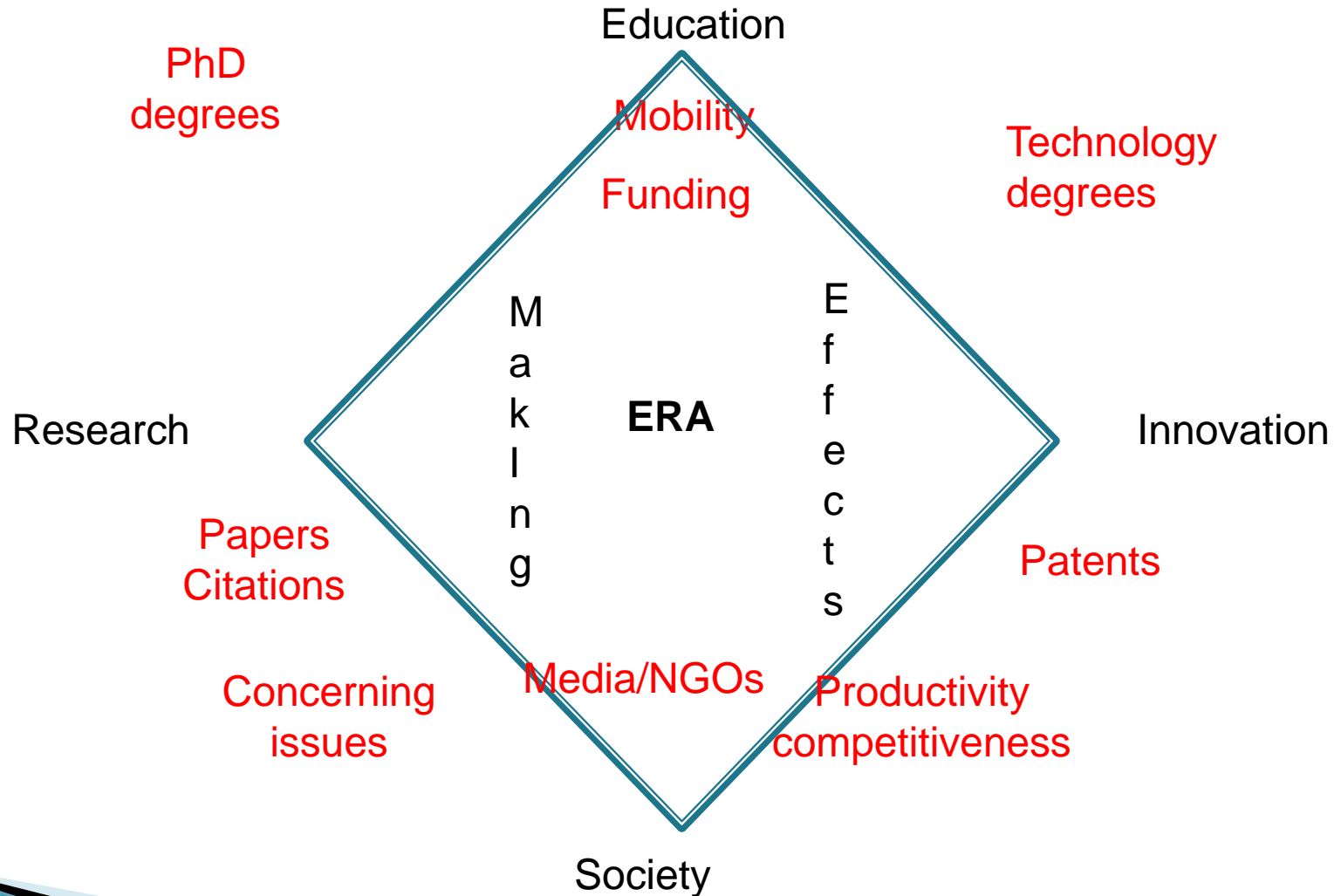
Competitiveness Council on 2 December 2008

*“By 2020, all actors fully benefit from the ‘Fifth Freedom’ across the ERA: free circulation of researchers, knowledge and technology. The ERA provides attractive conditions and effective and efficient governance for doing research and investing in R&D intensive sectors in Europe. It creates strong added value by fostering a healthy Europe-wide scientific competition whilst ensuring the appropriate level of cooperation and coordination. It is responsive to the needs and ambitions of citizens and effectively contributes to the sustainable development and competitiveness of Europe.”*

# Preliminary issues

- ▶ Fifth Freedom is about flows
  - Collaboration is about knowledge flows
- ▶ Composite Fifth Freedom indicators
  - Combining flow indicators in the Knowledge Diamond: Education, Research, Innovation, Society
- ▶ Cybermetrics (Webometrics) Tools
  - A new approach with some advantages
    - Universal (more stakeholders involved, global impact)
    - Powerful (Link analysis) and trusted method
    - Feasible (specially for large EU scenarios)
    - Impact (evaluation, monitoring, policies)

# Knowledge Diamond

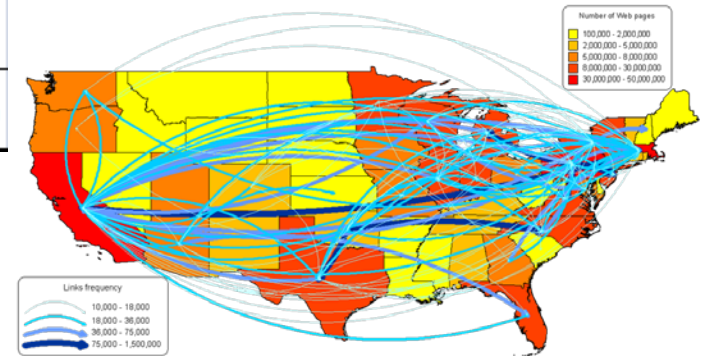
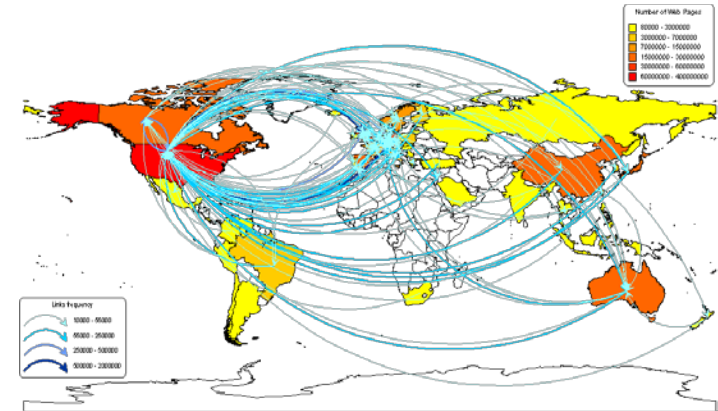


# Cybermetrics

- ▶ Quantitative analysis of academic presence and scholarly communication
  - Cyberscientometrics
- ▶ Two big areas
  - Flows of information through Internet networks
  - Flows of knowledge through web networks
- ▶ 3 families of indicators
  - Activity (web presence) related
  - Visibility (web impact) related
  - Usage related

# Flows

Flows	Indicator	Web Tools	Target
Information Flows	Internet Traffic	GRID Networks	Big Science
		Academic P2P Networks	e-Science
Knowledge Flows	Co-authoring	Google Scholar (Publish or Perish)	Research
	Bibliographic Coupling		
	Co-citation		
	Interlinking	Robots Search Engines	Education Research Innovation Society
	Co-linking		
	Wepage coupling		
	Web mention	Search Engines	
	WebTWOmetrics	Specialised Search Engines	Education, Research Society
	(Web) Usage	Logs from Repositories, e-Journals	Education, Research



# Webometrics

## ► Tools

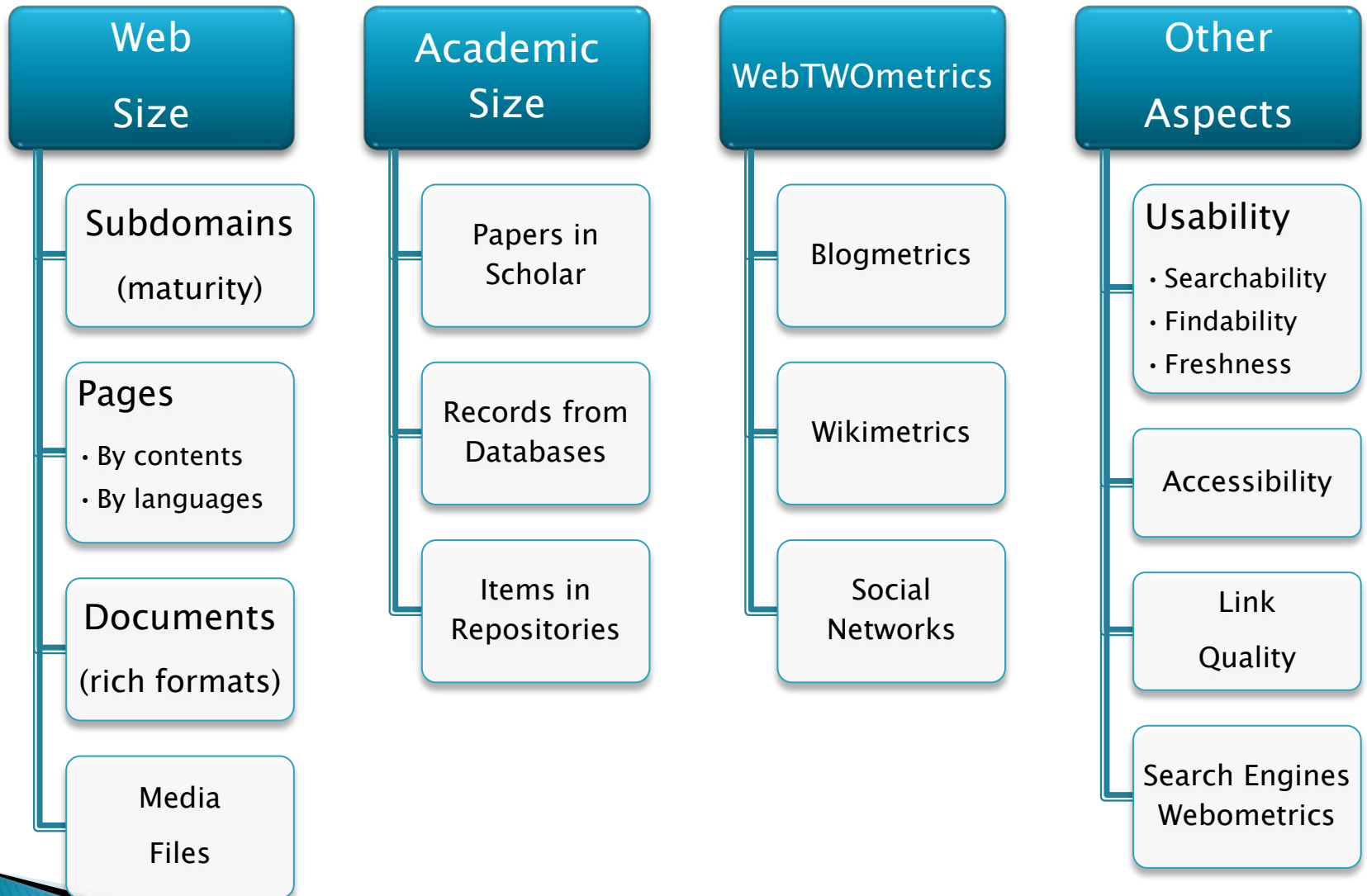
- Custom designed personal robots
  - Resources, time consuming
- (Commercial) Search engines
  - Visibility intermediaries

## ► Units

- Web domain/subdomain is the “natural” unit
  - Few duplicate, but changes–mergers–splits
- Individual files (pdf, doc, video, ...)
- Name of author/institution, title of a paper, patent, book
  - More problematic because of variants



# Activity (web presence)





# Ranking Web

July 09

## Ranking Web of European Universities

**Data**

- About Us
- About the Ranking
- Top 500 European Univ.
- Top 100 by Country
- Best Practices
- Notes

**Catalogue**

- Universities by country

**Information**

- Methodology
- Glossary
- Blog
- Links
- Contact Us/Disclaimer
- Site Map

Search Webometrics for:


Submit

Category:

All v

Top 500 European Universities								
RANKING		UNIVERSITY	COUNTRY	SIZE	VISIB.	POSITION		
EUROPE	WORLD					RICH FILES	SCHOLAR	
1	22	University of Cambridge		26	13	70	86	
2	42	University of Oxford		122	19	51	157	
3	46	Swiss Federal Institute of Technology ETH Zurich		84	49	76	43	
4	51	University College London		100	39	59	126	
5	52	University of Helsinki **		67	89	43	56	
6	54	Norwegian University of Science & Technology		88	58	86	45	
7	55	University of Oslo		62	67	39	122	
8	68	Universität Wien		130	87	57	67	
9	69	Universidad Complutense de Madrid		108	110	185	5	
10	71	University of Edinburgh		196	59	63	124	
11	72	Università di Bologna		126	66	110	109	
12	74	University of Southampton		383	88	104	19	
13	79	Utrecht University		171	121	77	63	
14	80	Uppsala University		90	147	82	70	
15	81	Université Catholique de Louvain *		63	178	247	10	
16	84	Freie Universität Berlin		74	85	228	106	
17	88	Université de Geneve		39	125	105	257	
18	92	Linköping University		66	188	141	68	
19	94	University of Groningen		58	207	164	59	
20	97	Charles University		150	90	75	306	

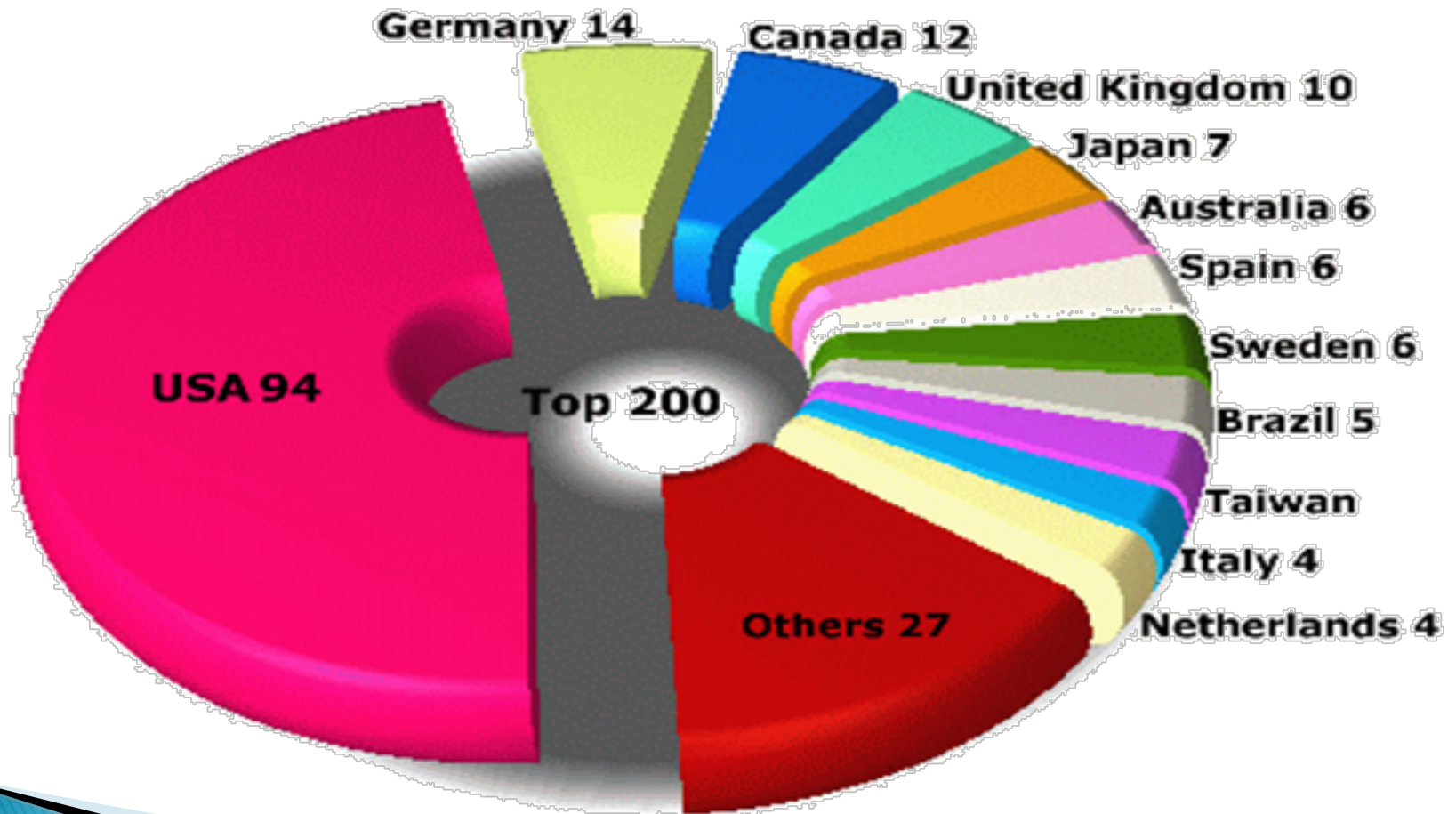
**CybermetricsLab**



www.mavir.net

# Academic Digital Divide

**Universities Distribution by Country**



(July, 2009)

# Other rankings

Top 200	ARWU09	WRJu109	Heeact09	THE09	Top 200	ARWU09	WRJu109	Heeact09	THE09
USA	90	94	87	54	UK	23	10	19	29
Canada	6	12	9	11	Germany	14	14	18	10
Japan	9	7	8	11	Netherlands	9	4	8	11
Australia	6	6	6	9	Switzerland	6	3	7	7
Israel	4	1	3	3	Sweden	4	6	4	5
China	0	2	3	6	France	7	1	5	4
Taiwan	1	5	1	1	Italy	4	4	7	1
Korea	1	1	2	4	Belgium	4	1	3	5
Hong Kong	0	2	1	5	Spain	1	6	2	1
Brazil	1	5	1		Denmark	3	1	2	3
Singapore	1	1	1	2	Norway	1	3	1	2
Mexico	1	1		1	Austria	1	2	1	1
New Zealand				3	Finland	1	1	1	1
Thailand		1		1	Russia	1	1		2
India				2	Ireland		1		2
Argentina	1				Czech Rep		2		
Saudi Arabia		1			Portugal		1		
South Africa				1	Greece				1
Malaysia				1	<i>countries</i>	<i>25</i>	<i>31</i>	<i>24</i>	<i>32</i>

# Impact

## Web Mention (Invocation)

### Term/phrase

- Frequency
- Positioning (Ranking)

### URLs

### Academics

- Name/ Email
- Institution's Name
- Title of Paper

## Link Visibility

### No. Inlinks

- G-factor
- by Domain

### Weighted

- PageRank

### Motivations

- "sitations"
- shallow/deep linking
- Content analysis

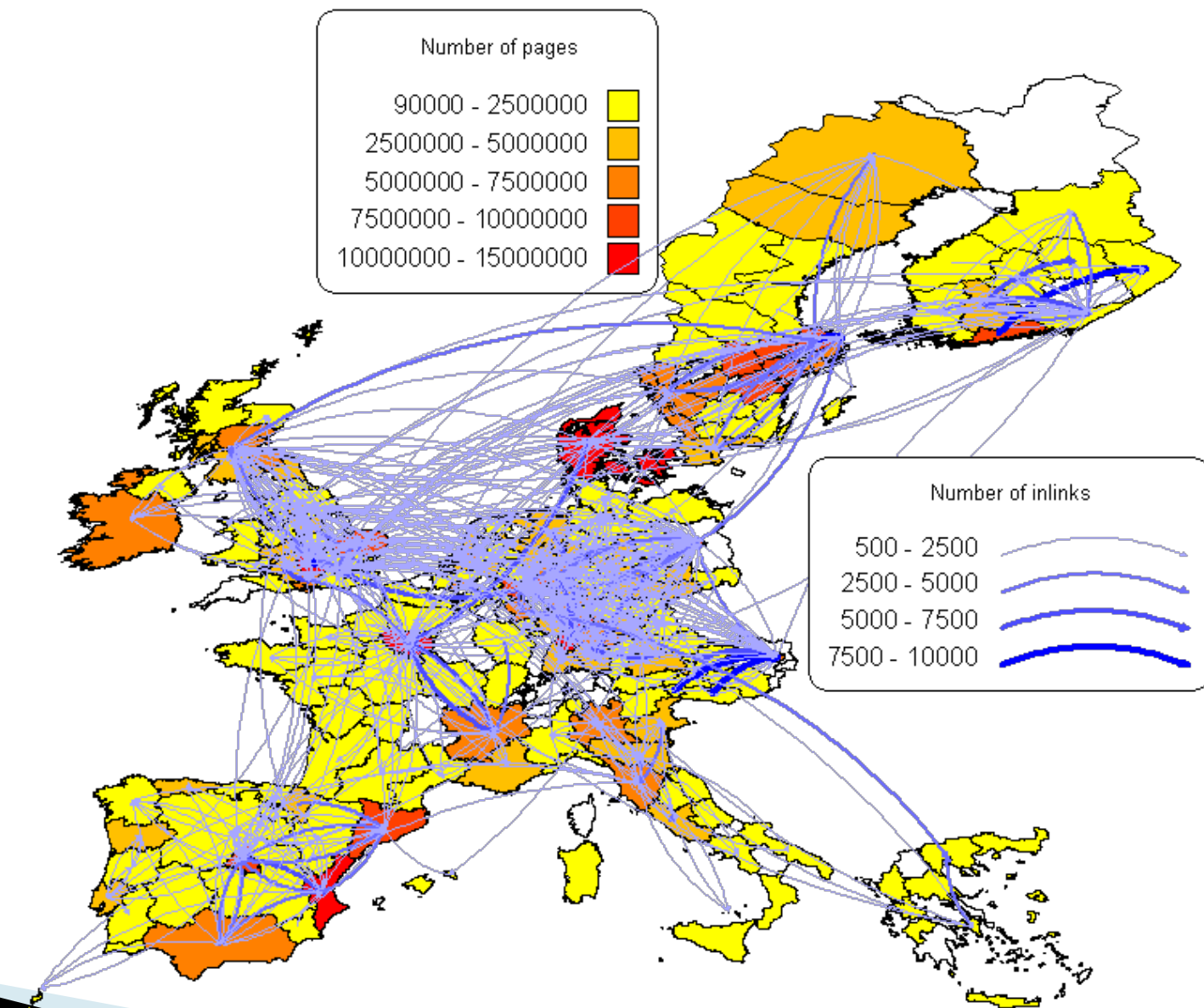
## Network Analysis

### Inter- linking Co- link

### Neighborhood (similarity)

- by Links
- by Terms
- by Visits

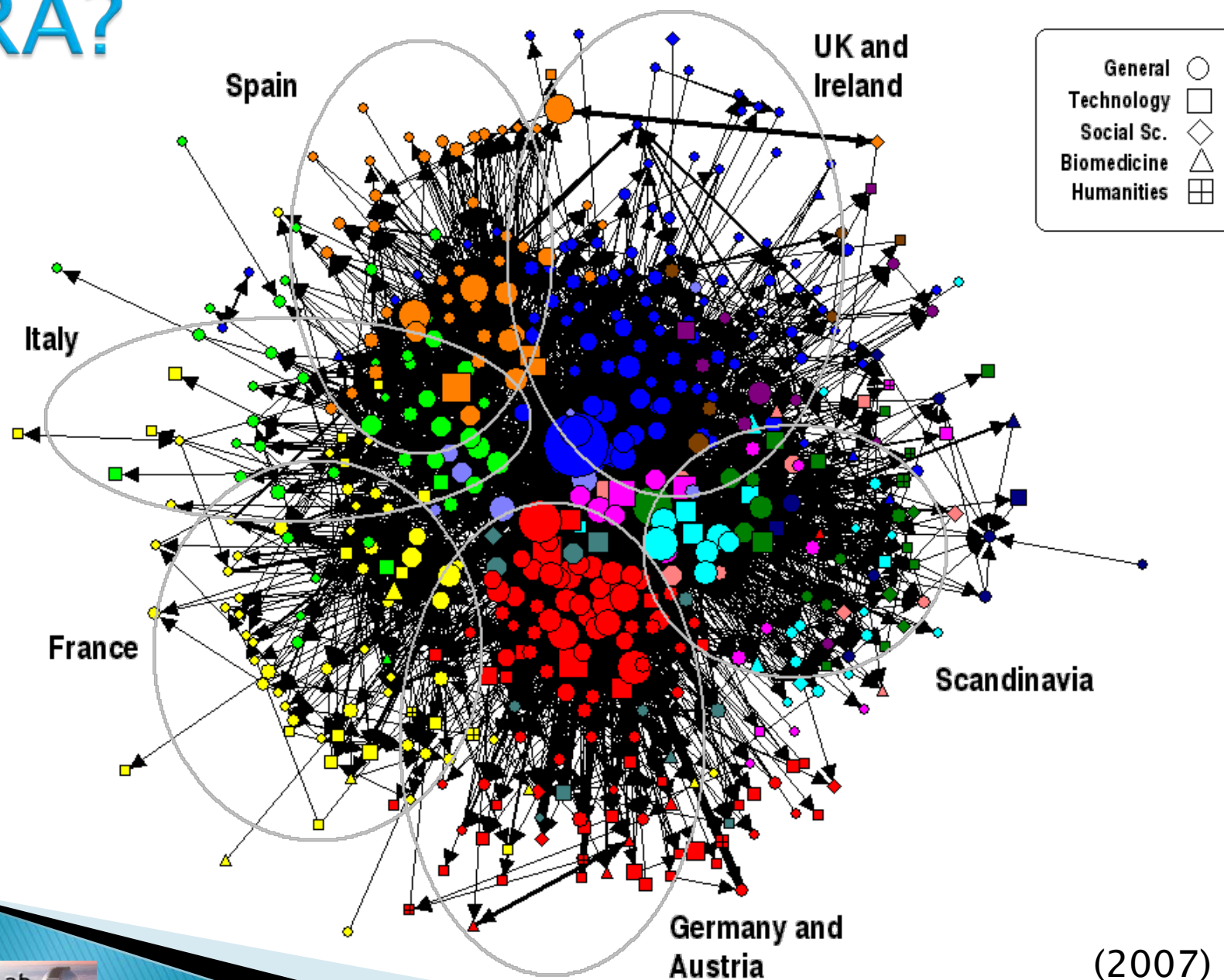
# ERA



(2007)



# ERA?



(2007)

# Link Analysis

- ▶ Interlinking and co-linking
  - +500 Universities & Research Centers
    - at least 10 for each 35 European countries (incl. Turkey & Israel)
  - 50 Top universities (North America)
  - East Asia (10 universities each from Japan, Korea, Taiwan)
  - BRIC (10 each from Brazil, Russia, India and China-Hong Kong)
- ▶ Feasibility
  - $2*600*600 = 0.72$  million pairs (requests)



# Web Mention Analysis

- ▶ Hot Topics (incl. Grand Challenges)
  - 50 individual terms or short sentences
  - 24 languages of the IATE terminological database
- ▶ Mentioning websites
  - 100 Top media
  - 100 NGOs
  - 100 academia
  - 100 governmental units
- ▶ Feasibility
  - $50 * 24 * 400 = 0.48$  million pairs (requests)

# Combined Individual Analysis

- ▶ Bibliometric annotation
  - 10 Highly Cited Researchers \* 35 countries
    - 10 most cited publications (5 last years)
    - Grouping by field, institutions, gender
- ▶ Mentioning websites
  - Search Engines
    - From academic (45) and non-academic domains (5)
  - Specialized Search Engines: Google Scholar
    - Mentions/Citations
- ▶ Feasibility
  - $3500 * 3 * 51 = 0.48$  million pairs (requests)

# Usage (Web metrics)

## Popularity

### Visits

- TrafficRank
- Pageviews

### Distribution

- by page/file
- by directory

### Visitors

- Geography
- Temporal distribution
- Robots

## Behaviour

### Entry/Exit Pages

- Time spent
- Bounce rate
- Repeated visits

### Referrals

- Search engines
- Referrers

## Downloads

### Papers

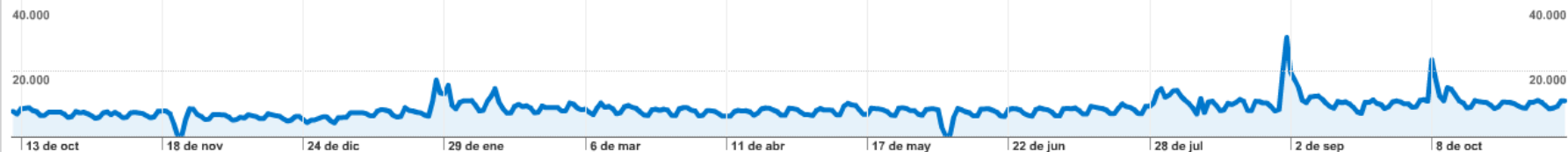
- by author
- by institution
- by journal
- by discipline

## Panel

11/10/2008 - 11/11/2009

Visitas

Gráfico realizado por:



### Uso del sitio

**3.374.815** Visitas

**19.620.692** Páginas vistas

**5,81** Páginas/visita

**38,08%** Porcentaje de rebote

**00:05:26** Promedio de tiempo en el sitio

**70,77%** Porcentaje de visitas nuevas

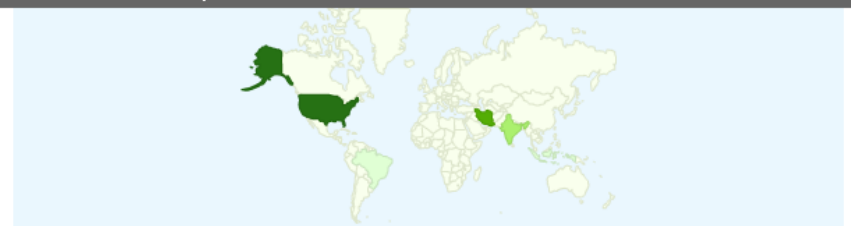
### Visión general de usuarios



**2.419.389** Usuarios del sitio web

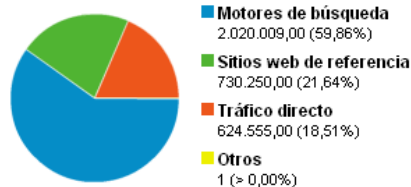
[ver informe](#)

### Gráfico de visitas por ubicación



[ver informe](#)

### Visión general de las fuentes de tráfico



[ver informe](#)

### Visión general del contenido

Páginas	Páginas vistas	Porcentaje de página
/	1.410.614	7,54%
/top4000.asp	671.009	3,59%
/top6000.asp	631.226	3,38%
/rank_by_country_select.asp	525.097	2,81%
/top100_continent.asp?cont=europe	515.405	2,76%

[ver informe](#)

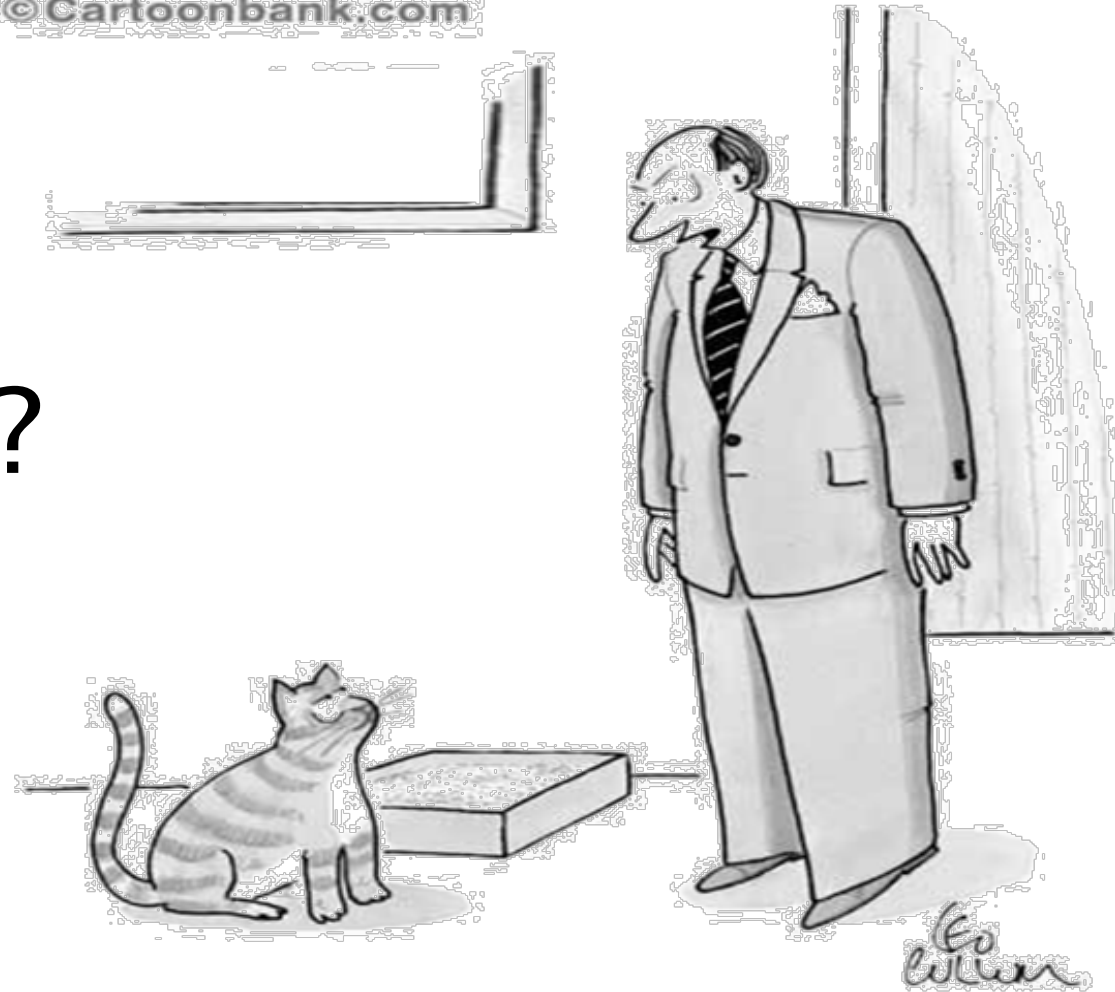
# Final comments

- ▶ Web indicators are useful monitoring ERA
  - Especially topics related to Fifth Freedom and information/knowledge flows
- ▶ Web networks can be built using different sources and techniques
  - Including new and innovative methods
  - Being feasible even for large populations
- ▶ Usage data analysis very promising
  - But still lot of methodological problems

# Thank you!

© Cartoonbank.com

► Questions?



*"Never, ever, think outside the box."*