ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

ENLM

E.Guevara & S.Scalise

A Proposal for a Common European Bibliographical Source

with the collaboration of L. Ducceschi and E. Zanchetta

Target and Limits of this talk

- Target: to immagine an European tool for bibliographical information (and maybe more…), free, interactive, and with less shorthcomings than the existing ones
- Limits: all examples are from only one field (linguistics) and linguistics is in many senses just halfway between humanities and 'hard' sciences.

Existing Citation Databases

- Openacademia http://www.openacademia.org

  An open source publication for scientific communities (not working at the moment). It will contain data only of registered people.

- Scopus http://www.scopus.com

  "The largest abstract and citation database of research literature and quality web sources covering over 16,000 peer-reviewed journals from more than 4,000 publishers". (24 entries for Noam Chomsky)

- Linguistic Bibliography Online

  It doesn't have any exportation capability. It covers the language niche areas that Brill specializes in.

Existing Citation Databases

- JSTOR Archives of over one thousand leading academic journals across the humanities, social sciences, and sciences, as well as select monographs and other materials valuable for academic work.
  - only 3 journals of linguistics; no books
  - incomplete exportations capabilities

- Google Scholar offers the possibility of searching a large web based database of citation and references and to export each result as a BibTeX entry. Free.
  - A lot of 'noise'
  - Reduced Exportation capabilities (one record at the time)

Existing Citation databases

- ISI Web of Knowledge – Thomson Reuters

http://isiknowledge.com/

Relevance of Citation Databases

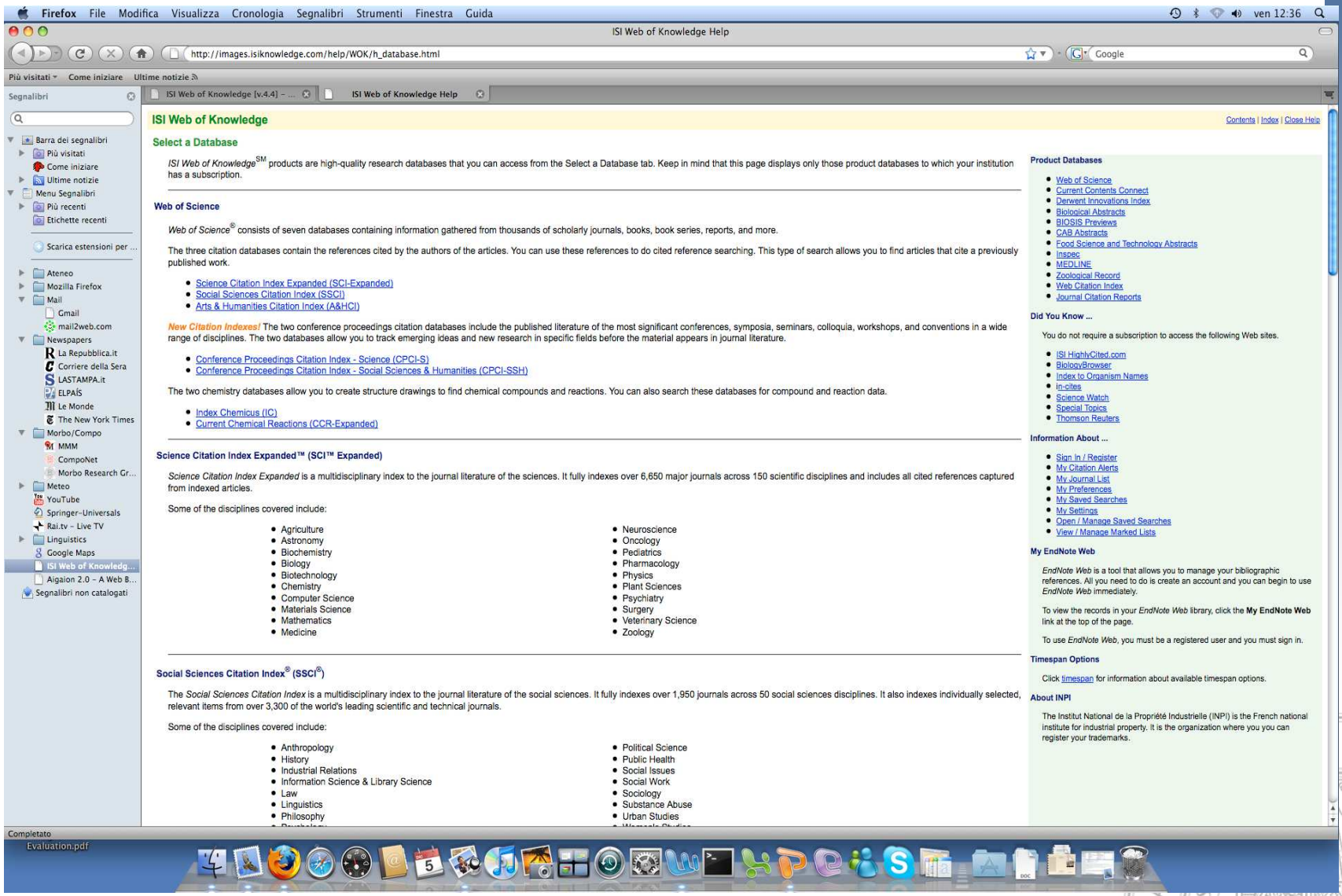
- DLR (German Aerospace Centre) Institute for Communications and Navigation

  Deutsches Zentrum für Luft- und Raumfahrt e.V. - München – Germany

- Parts of the funding depends on the publications of the staff.

  But only publications quoted on *Web of Knowledge* count

ISI Web of Knowledge – Thomson Reuters



ISI Web of Knowledge – Thomson Reuters

- Positive features of Thomson of knowledge according to Henk Moed

→

ISI web of Knowledge – Thomson Reuters

Unique properties of Thomson WoS (H.Moed)

- Has multi-disciplinary coverage
- Processes leading, international journals
- Processes journals cover-to-cover
- Includes cited references
- Includes all authors
- Includes all authors' institutional affiliations

ISI Web of Knowledge – Thomson Reuters

- other interesting features:

- based on very large databases (but…)
- keywords
- ‘refine results’
- exportation capabilities (but…)

ISI Web of Knowledge – Thomson Reuters: limitations

- Criteria for the selection of Journals (cf. the non inclusion of "Linguistic Typology")
- Very few monographic volumes (e.g. there is no *Foundations of Language* by R. Jackendoff:

  "Web of Science consists of seven databases containing information gathered from thousends of scholarly journals, books, book series, reports and more."

  (there are reviews of fundamental books but not the books themselves)
- Time lapse before a publication is indexed (cf. "Lingue e Linguaggio": two years and yet …)
- It seems more dedicated to the 'hard sciences' (where a typical product is a short paper in refereed journals)

ISI Web of Knowledge – Thomson Reuters: limitations

- Exportation: the tool is present but exportation in .bib extension for BibTeX application (which is massively used by researchers in 'hard' sciences) is rather omplicated and sometimes gives wrong results.
- Humanities are not well represented

  (H.Moed: 56% of A&HCI documents has no address)
- Not free, but subscription-based

ISI web of Knowledge – Thomson Reuters: limitations

- Example 1: Noam Chomsky
  - 9 pages, but … most of the items are about politics and not linguistics

1. Title: An open letter on Iran Barack Obama
Author(s): Aronowitz, S; Benjamin, M; Chomsky, N, et al.
Source: **NEW YORK REVIEW OF BOOKS** Volume: **55** Issue: **13** Pages: **73-73** Published: **2008**
Times Cited: **0**
A-link

2. Title: 'Against us military bases in the Czech Republic'
Author(s): Landy, J; Harrison, T; Shalom, SR, et al.
Source: **NEW YORK REVIEW OF BOOKS** Volume: **55** Issue: **2** Pages: **58-58** Published: **2008**
Times Cited: **0**
A-link

3. Title: Statement by scholars of Iran, the Middle East, and others protesting the detention of Dr. Haleh Esfandiari by the Iranian government
Author(s): Abrahamian, E; Afary, J; Afkhami, M, et al.
Source: **NEW YORK REVIEW OF BOOKS** Volume: **54** Issue: **11** Pages: **8-9** Published: **JUN 28 2007**
Times Cited: **0**
A-link

4. Title: Homi Bhabha talks with Noam Chomsky
Author(s): Bhabha, H; Chomsky, N
Source: **CRITICAL INQUIRY** Volume: **31** Issue: **2** Pages: **419-424** Published: **WIN 2005**
Times Cited: **0**
A-link

5. Title: Targeted in Lebanon
Author(s): McCarthy, T; Balaghi, S; Klein, N, et al.
Source: **NEW YORK REVIEW OF BOOKS** Volume: **52** Issue: **19** Pages: **69-69** Published: **DEC 1 2005**
Times Cited: **0**
A-link

6. Title: Made in America - A century of US intervention and subversion culminates in Haiti's latest coup
Author(s): Chomsky, N
Source: **INDEX ON CENSORSHIP** Volume: **33** Issue: **2** Pages: **18-25** Published: **APR 2004**
Times Cited: **0**
A-link

7. Title: Chomsky's revolution (Comment)
Author(s): Chomsky, N
Source: **NEW YORK REVIEW OF BOOKS** Volume: **49** Issue: **12** Pages: **64-64** Published: **JUL 18 2002**
Times Cited: **0**
A-link

8. Title: Confronting the monster - The US has a long unsavoury history in its fight against terror
Author(s): Chomsky, N
Source: **INDEX ON CENSORSHIP** Volume: **31** Issue: **1** Pages: **24-29** Published: **JAN 2002**
Times Cited: **0**
A-link

9. Title: An open letter to members of the Board of the Pacifica and Mary Frances Berry, Chair
Author(s): Walker, A; Davies, A; Baez, J, et al.
Source: **INDEX ON CENSORSHIP** Volume: **28** Issue: **5** Pages: **147-147** Published: **SEP-OCT 1999**
Times Cited: **0**
A-link

10. Title: Into the future (The successes and failures of the Universal Declaration of Human Rights)
Author(s): Udhari, A; Neier, A; Zhirinovsky, V, et al.
Source: **INDEX ON CENSORSHIP** Volume: **27** Issue: **3** Pages: **184-189** Published: **MAY-JUN 1998**
Times Cited: **0**
A-link

11. Title: Serial veto (The Palestinian Israeli conflict and UN resolution 242)
Author(s): Chomsky, N
Source: **INDEX ON CENSORSHIP** Volume: **26** Issue: **6** Pages: **155-161** Published: **NOV-DEC 1997**
Times Cited: **0**
A-link

12. Title: Democracy lite + The 1996 US electoral campaign
Author(s): Chomsky, N
Source: **INDEX ON CENSORSHIP** Volume: **26** Issue: **3** Pages: **56-58** Published: **MAY-JUN 1997**
Times Cited: **0**
A-link

13. Title: LANGUAGE AND EVOLUTION - REPLY
Author(s): CHOMSKY, N
Source: **NEW YORK REVIEW OF BOOKS** Volume: **43** Issue: **2** Pages: **41-41** Published: **FEB 1 1996**
Times Cited: **0**
A-link

14. Title: NO ENTRY + IN RESPONSE TO HARRIS,ROY
Author(s): CHOMSKY, N
Source: **TLS-THE TIMES LITERARY SUPPLEMENT** Issue: **4804** Pages: **19-19** Published: **APR 28 1995**
Times Cited: **0**
A-link

15. Title: AMERICAN FOREIGN-POLICY - REPLY
Author(s): CHOMSKY, N
Source: **TLS-THE TIMES LITERARY SUPPLEMENT** Issue: **4796** Pages: **15-15** Published: **MAR 3 1995**
Times Cited: **0**
A-link

16. Title: SWEET HOME OF LIBERTY + HUMAN-RIGHTS VIOLATIONS IN THE UNITED-STATES
Author(s): CHOMSKY, N
Source: **INDEX ON CENSORSHIP** Volume: **23** Issue: **3** Pages: **9-18** Published: **JUL-AUG 1994**
Times Cited: **0**
A-link

17. Title: UNIVERSAL GRAMMAR
Author(s): CHOMSKY, N
Source: **NEW YORK REVIEW OF BOOKS** Volume: **38** Issue: **21** Pages: **82-82** Published: **DEC 19 1991**
Times Cited: **0**
A-link

18. Title: AFTER THE COLD-WAR, UNITED-STATES-FOREIGN-POLICY IN THE MIDDLE-EAST
Author(s): CHOMSKY, N
Source: **CULTURAL CRITIQUE** Issue: **19** Pages: **15-31** Published: **FAL 1991**
Times Cited: **0**
A-link

- if one refines the research -> Linguistics, the result are five publications:

- (a funny result for the most influential linguist in the world…)

ALL DATABASES

<< Back to previous

Results Author=(Chomsky N.)
Timespan=All Years.
Refined by: Subject Areas=(LINGUISTICS)

Results: **5** Page 1 of 1 Go

Print | E-mail | Add to Marked List | Save to EndNote Web | Save to EndNote, RefMan, ProCite | more options

Refine Results

Search within results for
Search

General Categories Refine
- SOCIAL SCIENCES (5)
- SCIENCE & TECHNOLOGY (1)

more options / values...

Subject Areas Refine
- LINGUISTICS (5)
- BEHAVIORAL SCIENCES (1)
- EVOLUTIONARY BIOLOGY (1)
- GENETICS & HEREDITY (1)
- MATHEMATICS (1)

more options / values...

Document Types

Authors

Source Titles

Publication Years

Languages

For advanced refine options, use
Analyze Results

1. Title: Message from Noam Chomsky
Author(s): Chomsky, N
Source: **LINGUA** Volume: **116** Issue: **10** Pages: **1469-1469** Published: **OCT 2006**
Times Cited: **0**
A-link

2. Title: Three factors in language design
Author(s): Chomsky, N
Source: **LINGUISTIC INQUIRY** Volume: **36** Issue: **1** Pages: **1-22** Published: **2005**
Times Cited: **39**
A-link

3. Title: The evolution of the language faculty: Clarifications and implications
Author(s): Fitch, WT; Hauser, MD; Chomsky, N
Source: **COGNITION** Volume: **97** Issue: **2** Pages: **179-210** Published: **2005**
Times Cited: **28**
A-link

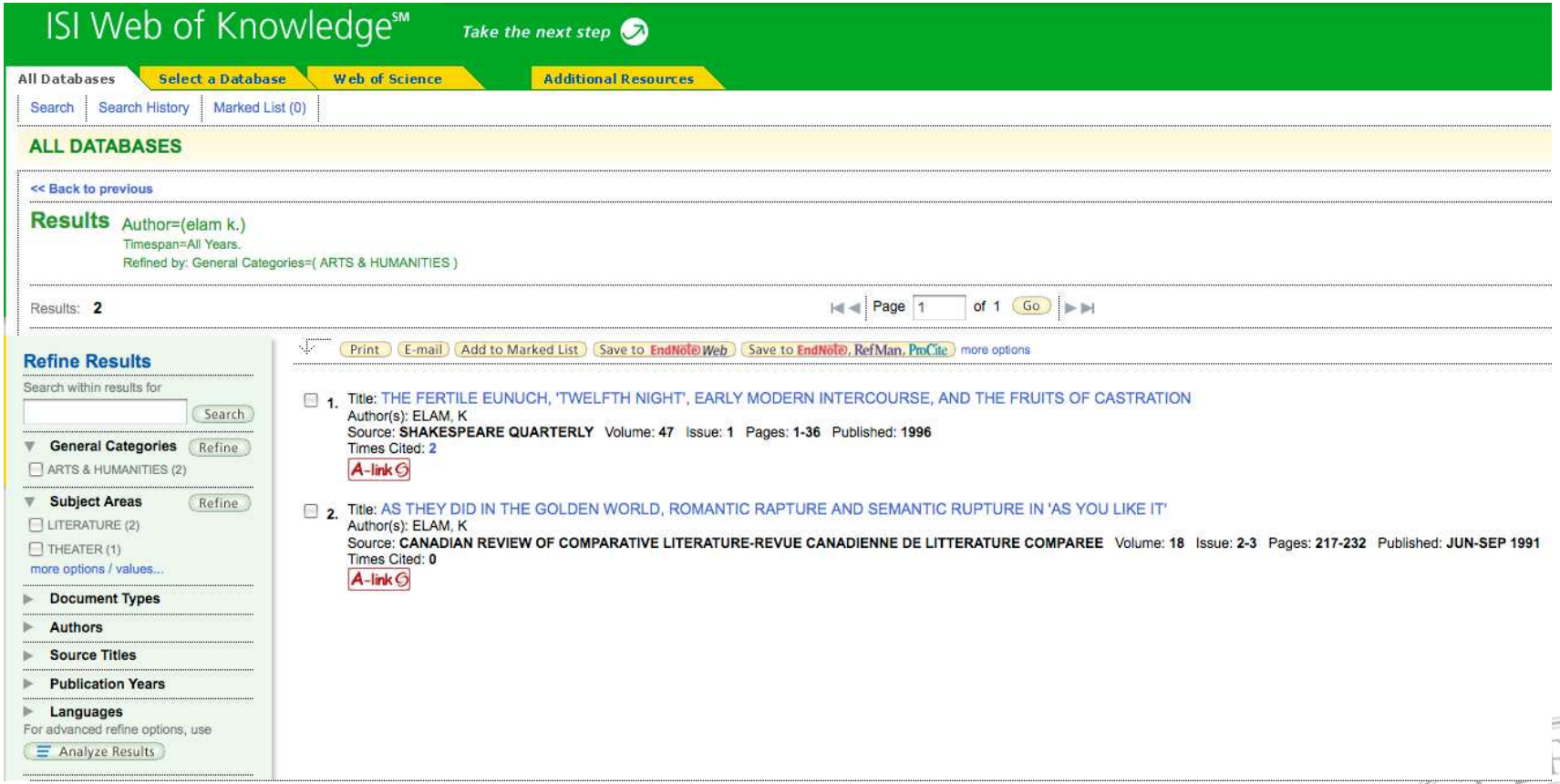
4. Title: The faculty of language: What is it, who has it, and how did it evolve?
Author(s): Hauser, MD; Chomsky, N; Fitch, WT
Source: **SCIENCE** Volume: **298** Issue: **5598** Pages: **1569-1579** Published: **NOV 22 2002**
Times Cited: **382**
A-link

5. Title: TOPIC ... COMMENT - ON FORMALIZATION AND FORMAL LINGUISTICS
Author(s): CHOMSKY, N
Source: **NATURAL LANGUAGE & LINGUISTIC THEORY** Volume: **8** Issue: **1** Pages: **143-147** Published: **FEB 1990**
Times Cited: **5**
A-link

Results: **5** Show 10 per page Page 1 of 1 Go

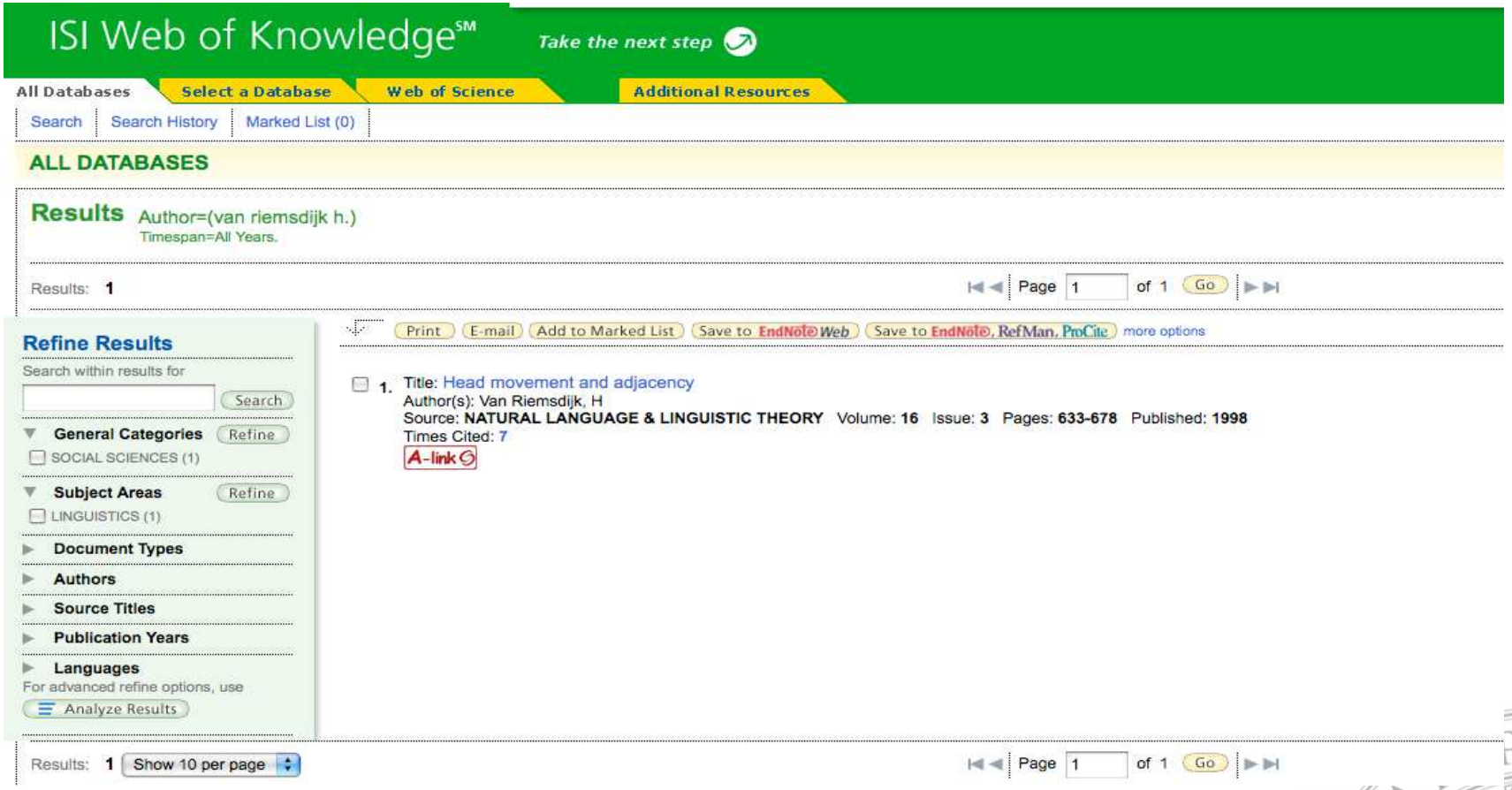
- on the other end, the director of the department of Foreign Languages and co-organizer of this conference, Keir Elam is quoted two times:

Example 2


ISI Web of Knowledge℠
Take the next step
All Databases
Select a Database
Web of Science
Additional Resources
Search
Search History
Marked List (0)
ALL DATABASES
<< Back to previous
Results Author=(elam k.)
Timespan=All Years.
Refined by: General Categories=(ARTS & HUMANITIES)
Results: 2
Page 1 of 1 Go
Refine Results
Search within results for
Search
General Categories Refine
ARTS & HUMANITIES (2)
Subject Areas Refine
LITERATURE (2)
THEATER (1)
more options / values...
Document Types
Authors
Source Titles
Publication Years
Languages
For advanced refine options, use
Analyze Results
Print E-mail Add to Marked List Save to EndNote Web Save to EndNote, RefMan, PmCite more options
1. Title: THE FERTILE EUNUCH, 'TWELFTH NIGHT', EARLY MODERN INTERCOURSE, AND THE FRUITS OF CASTRATION
Author(s): ELAM, K
Source: SHAKESPEARE QUARTERLY Volume: 47 Issue: 1 Pages: 1-36 Published: 1996
Times Cited: 2
A-link
2. Title: AS THEY DID IN THE GOLDEN WORLD, ROMANTIC RAPTURE AND SEMANTIC RUPTURE IN 'AS YOU LIKE IT'
Author(s): ELAM, K
Source: CANADIAN REVIEW OF COMPARATIVE LITERATURE-REVUE CANADIENNE DE LITTERATURE COMPAREE Volume: 18 Issue: 2-3 Pages: 217-232 Published: JUN-SEP 1991
Times Cited: 0
A-link

- ..and our invited speaker, Henk van Riemsdijk, one of the'fathers' of Generative Linguistics in Europe and very well know in the whole world:

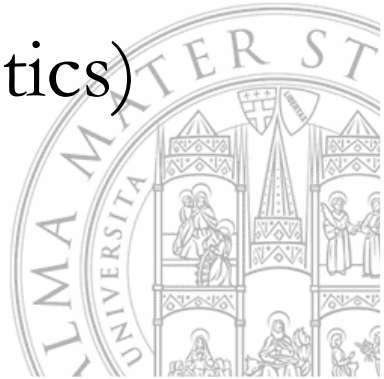
Example 3


ISI Web of Knowledge℠
Take the next step
All Databases
Select a Database
Web of Science
Additional Resources
Search
Search History
Marked List (0)
ALL DATABASES
Results
Author=(van riemsdijk h.)
Timespan=All Years.
Results: 1
Page 1 of 1 Go
Refine Results
Search within results for
Search
General Categories Refine
SOCIAL SCIENCES (1)
Subject Areas Refine
LINGUISTICS (1)
Document Types
Authors
Source Titles
Publication Years
Languages
For advanced refine options, use
Analyze Results
Print E-mail Add to Marked List Save to EndNote Web Save to EndNote, RefMan, ProCite more options
1. Title: Head movement and adjacency
Author(s): Van Riemsdijk, H
Source: NATURAL LANGUAGE & LINGUISTIC THEORY Volume: 16 Issue: 3 Pages: 633-678 Published: 1998
Times Cited: 7
A-link
Results: 1 Show 10 per page
Page 1 of 1 Go

- Conclusion: WoK covers humanities quite badly

Are different approaches possible?

- Try not to rely *only* on private agencies / private content distributors
  - Distributed content input (the Wikipedia philosophy!)
- Make content sources free
  - Free to view
  - Free to update
- Improve coverage and completeness of data resources (critical for humanities - linguistics)
- Keep up to date



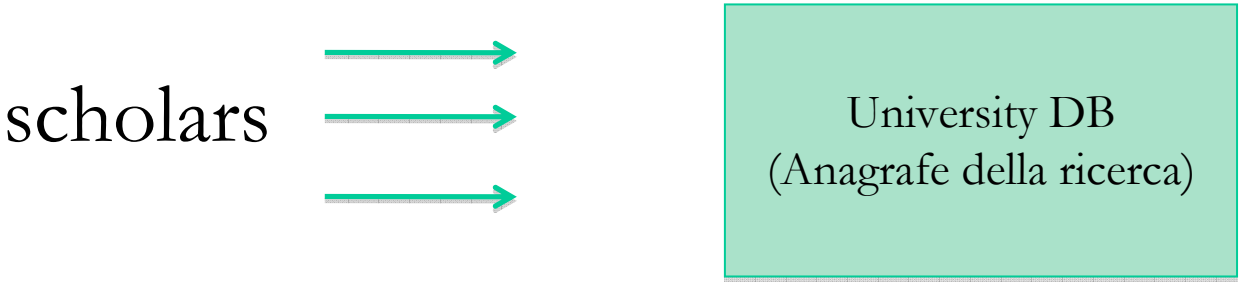
Where does the bibliographical info go today?

- Many Universities and Ministery agencies keep records for each research product / publication in dedicated databases
- These are primarily bibliographical data
- These are potentially public data
- Compiled by researchers themselves

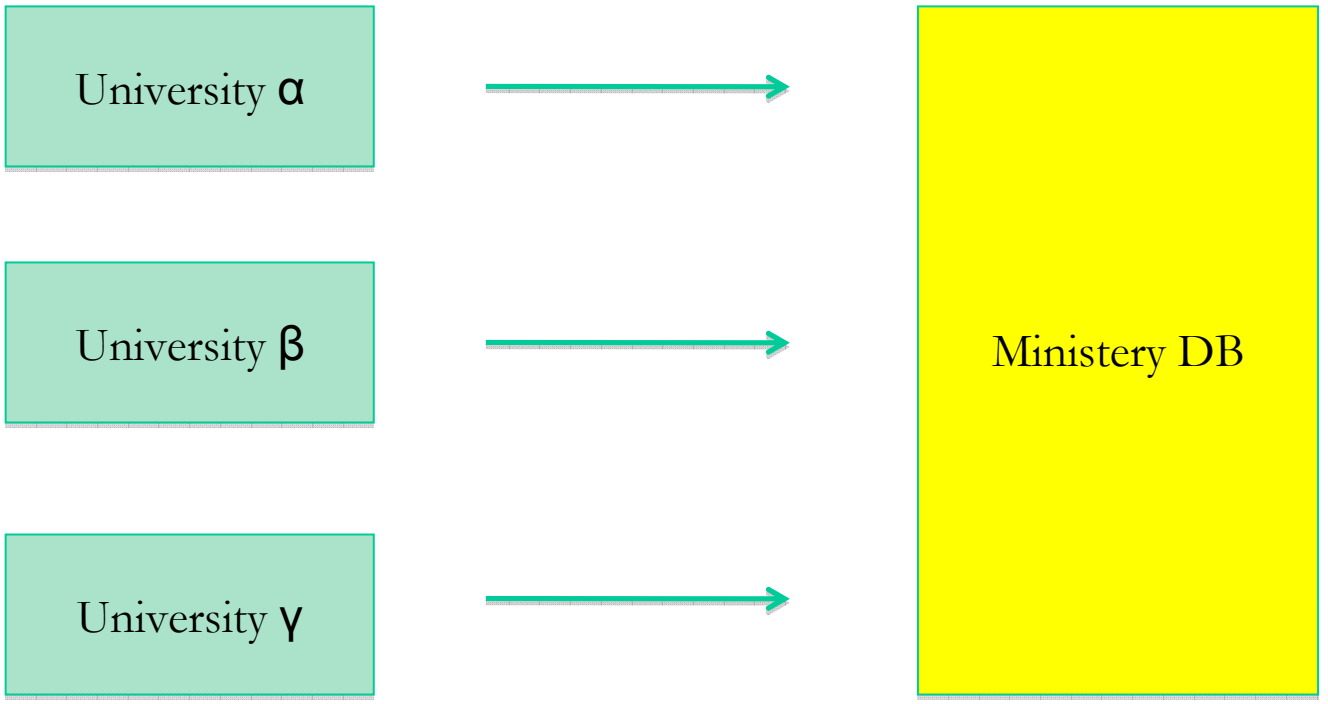
Where does the bibliographical info go today?

- Problems:
  - researchers fill the forms with their own scientific products but they cannot *use* them
  - data sparseness and non-uniform database formats
- Harmonize and align the different sources and make them available

Data flow: from researchers to Universities



Data flow: from Universities to the Ministery of the University

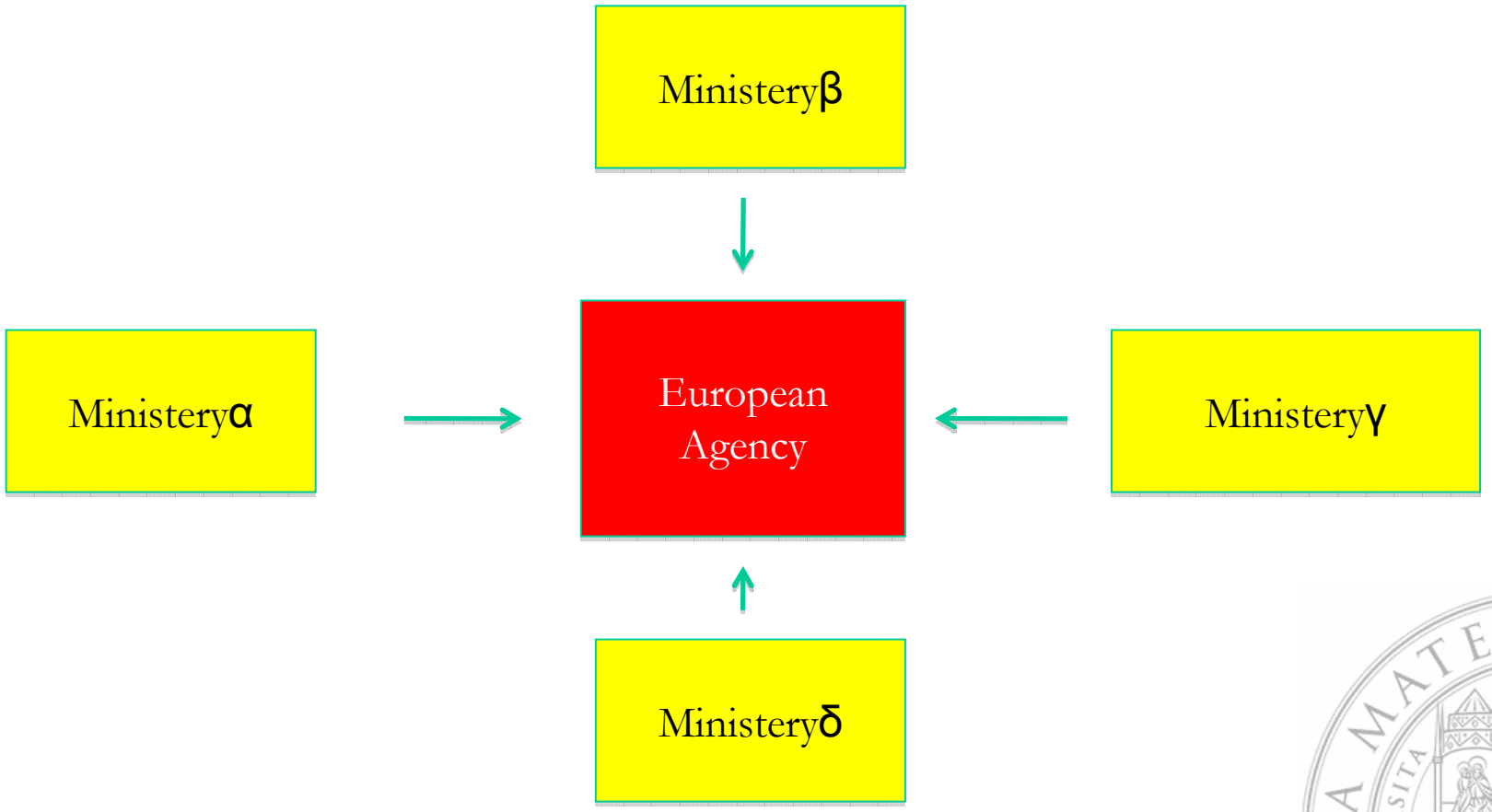


Data flow: from Universities to the Ministery of the University

- These data have NO general visibility.
- They are stored, probably accessible only to the researcher that compiled them, but not to the scientific community. They are good quality data, but silent and lost for general purposes.

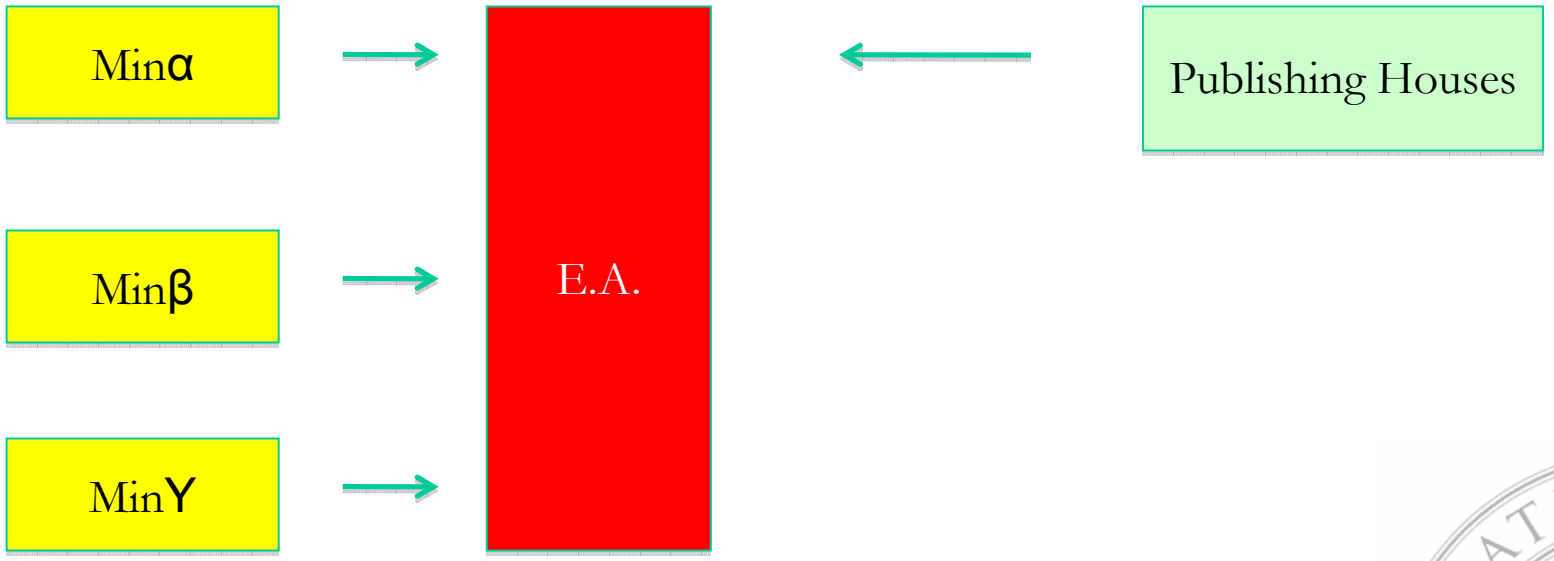
What should we do at the European level?

- Concentrate the distributed sources



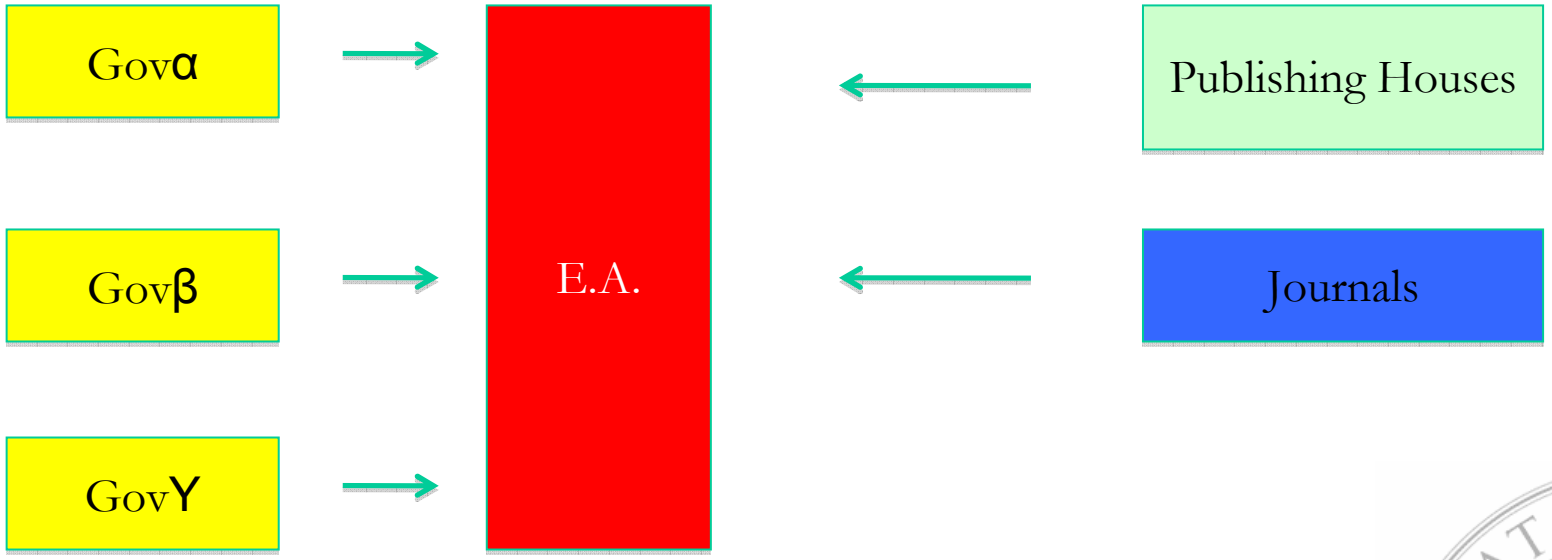
What else could we do at the European level?

- Get publishers involved
  (it's free advertising, after all!)



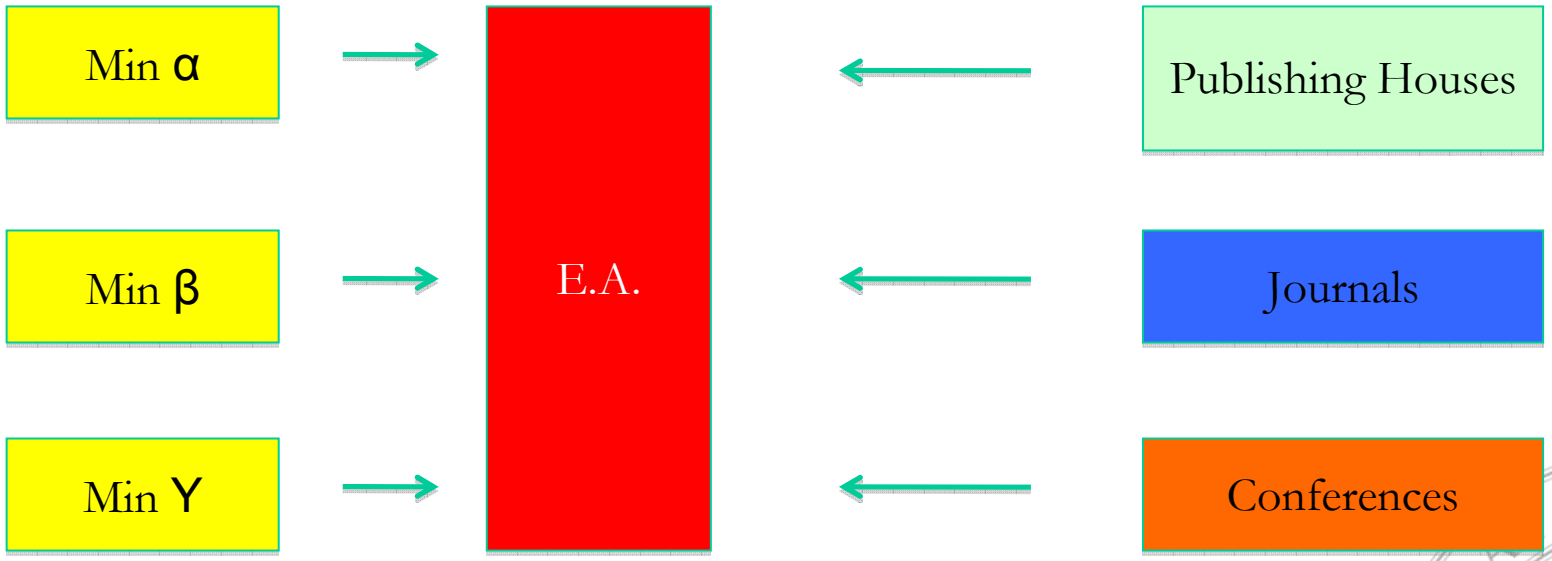
What else could we do at the European level?

- Index whole journals
  (let the publishers provide the data…)



What else could we do at the European level?

- Index conference proceedings and talks (let the organizers provide the data…)



Database maintenance and updates

- Not only data from Ministery agencies but also from:
  - Journals
  - Conferences
  - Publishers
- Compare and merge similar entries
- Let researchers do the quality control (distributed work…)

Our first trial: Biblio 1

- A bibliography of general linguistics
- Collected over many years of hard hand-work
  - First Biblio version (2004): ca. 8000 entries
  - http://morbo.lingue.unibo.it/biblio-lexical-morphology-a-bibliography.html

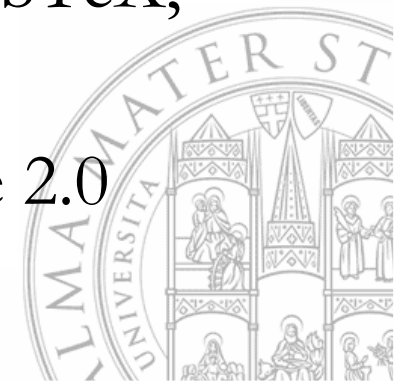
Bibliographic management: Aigaion

AIGAION: a small scale starting point is available

Aigaion is an open source application developped by a team of PhD Dutch students. We used their infrastructure and filled it with our data

- 25.000 entries, 23.000 authors
- It is a 'social network' application
- It can process many different bibliography database formats (= import and export BibTeX, Ris, etc.)

- Active development comunity (presently at the 2.0 development version)



Bibliographic management: Aigaion

- Web-based application (Php and Mysql)
- It's free
- It is not maintained by private enterprises
- Can be easily adapted and modified to suit different needs

Bibliographic management: Aigaion

- Not a definitive resource…
- Low budget and small work-force
- More a proof of concept than a real production site
- Such a wide ranging project would demand heavy computing infrastructure, but the technology is not too far from being ready
- Future EU cooperation projects could study the feasibility in detail

Bibliographic management: Aigaion

- Already now (in the field of linguistics)

  the results are maybe not exaustive but more representative (Chomsky and van Riemsdijk have some 60 publications listed: not enough but… a bit better)

- http://fraublucher.sslmit.unibo.it/aigaion/index.php/authors/show/23052

- http://fraublucher.sslmit.unibo.it/aigaion/index.php/authors/show/23055

Conclusions

An European tool such as the one sketchted here can become

- A tool for comprehensive bibliographical information
- A tool for measuring not only productivity of researchers but also a tool able to mesure the productivity of a whole department or of members of a specific research project.

Conclusions

- It can be useful to measure an imporatant parameter such as ‘continuity’
- It can be very easily used to automatically provide bibliographies of papers or books
- A bibliographical tool such as the one sketched here can easily integrate the ERIH A-B-C evaluation of Journals. It can thus couple qualitative and quantitative aspects of research production.

Conclusions

- In any event, there should be a common European action in order to coordinate and/or unify energies and proposals that at present go in the same direction but at random.
- A common target would be to implement free instruments, more complete than the existing ones, with friendly interacting properties, and full exportation capacities.