The European Research Papers Archive: Quality Filters in Electronic Publishing

Michael Nentwich

Electronic publishing is increasing in most academic disciplines, incorporating not just electronic journals but also online working-paper series, sites that offer conference papers, and private home pages offering preprints and offprints. They are a main feature of what may be called "cyberscience" (Nentwich 1999 [https://quod.lib.umich.edu/j/jep/3336451.0005.109 /--european-research-papers-archive-quality-filters?trgt=p_nentwich;view=fulltext#p_nentwich]) [1] [#note1] Other features of this cyberscience landscape include teleworks, virtual laboratories, the changing roles in knowledge generation, artificial intelligence, and computer modeling. This paper looks at the future of electronic publishing, taking as a starting point European-integration research, the study of how Europe is coming together politically, legally, economically, and socially.

Sidebar [https://quod.lib.umich.edu/j/jep/3336451.0005.102/--sidebar-how-erpa-functions?rgn=main;view=fulltext]

Although the trend started almost a decade ago, European-integration research has been slow to join the move to a broad range of electronic publication. While some publications have been on line for some years now, several new sources for scholarly work started only recently. [2] [#note2] At the moment there are some twenty online series in the broad field of European-integration research put together by research institutions or scholarly associations. These papers form a "series" in the sense that they are usually numbered and there is some rule or principle as to what may be published in the series. In European-integration research it is not uncommon for most papers in a series to be published again in peer-reviewed journals, sometimes in the same form, and other times after being reworked considerably.

The online landscape includes more than 210 papers in the five series that form the European Research Papers Archive; approximately 150 papers in other series (most of which recently came online); some conference-paper series [3] [#note3] ; almost all the European integration-research journals (only a fifth of which have the full text available on the Internet); various papers on researchers' personal Web sites, and at least twenty newsletters about European issues.

The ERPA Initiative

In November 1997 the four most prominent producers of online working papers in the field decided to coordinate their electronic publishing operations. Their aim was to provide "a common access point for the online working paper series of the participating institutions in order to help researchers in the field of European-integration studies searching the growing number of working papers now available in the Internet." The founding members of the network were:

- the Academy of European Law and the Robert Schuman Centre of the European University Institute [http://www.iue.it/] in Florence, Italy,
- the Harvard Jean Monnet Chair [http://www.jeanmonnetprogram.org/] in Boston, Massachusetts,
- the Max Planck Institute for the Study of Societies [http://www.mpi-fg-koeln.mpg.de/] in Cologne, Germany, and
- the European Integration online Papers [http://eiop.or.at/eiop/index.php/eiop] (EIoP), based in Vienna, Austria.

Coordinated by the editor of European Integration online Papers, the group formed the management board of what would become the European Research Papers Archive (ERPA). ERPA is supported by contributions — both financial and in-kind — from the participating institutions. The University of Economics and Business Administration in Vienna, which already hosts the EIoP, also agreed to host ERPA. The implementation, which involved considerable changes to the servers of the participating institutions, took place in late spring and summer of 1998. ERPA was finally launched on 8 September 1998. Two months later a fifth series, Advanced Research on the Europeanisation of the Nation-State [http://www.arena.uio.no/] (ARENA) in Oslo, Norway, joined the network.

European Research Papers Archive

The Archive started with about one hundred online papers, and the number had doubled by summer of 1999. On average, ERPA's search engine is used between 75 and 135 times each day — non-trivial use in a community of "Europeanists" estimated to be between 2,000 and 3,000 globally. [4] [#note4] Two other collections have applied for membership since ERPA was launched.

In establishing ERPA, the board had to deal with some policy issues brought about by the Internet and the technology it

makes possible. The same issues will be faced by other disciplines seeking to establish their own archives. What follows are strictly personal views. They are based on my experiences as a scholar, as the editor of an online series and as a participant in the discussions that preceded the creation of ERPA.

The Key Question

Any project aiming at building an archive for E-publications has to address one central question: Should the archive offer access to online papers regardless of their type and quality? In other words, what is the balance between comprehensiveness of the archive and the review process of the publications included? These two aims are conflicting, but not necessarily mutually exclusive.

Given the widespread perception that "there is too much to read anyway," any mechanism that reduces the amount of material to be digested seems highly welcome. It might, therefore, make sense not to include student papers, non-refereed papers, and grey literature such as conference papers. If those papers are good, we assume that they would be published in a quality series or journal. If there is too much to read, an archive that includes only quality papers will be more useful than one that does not check for quality.

But do we really have too much to read? While it is true that a search for papers on a rather general topic can turn up an overwhelming number of results, a scholar's focus on a narrow subject can produce a short and valuable hit list. To rule out a group of papers hampers the serious scholar from the benefit of finding "bad" papers that contain a gem that helps understanding — an accurate empirical description can be useful even though the author's conclusions may be suspect, for instance.

Disciplines vary, however, not only in their publishing cultures but also in what is expected by an archive. For instance, those who use and contribute to the well-known physics e-preprint archive [http://xxx.lanl.gov/] seem to prefer quick prepublication without review over the traditional system of peer review and long publication delays. The reasons are twofold: Registering a research result in the archives means putting a date stamp on the paper to assure that the ideas or results are attributed to the author or authors first. Also, data posted in the archive is immediately available for further research. The quality check comes later, after the e-prints are submitted to journals (journal submission often is contemporaneous with submission to the archive). However, it is doubtful that the same reasoning applies to other fields as well, because physics always had a strong tradition of exchanging preprints; moving to the e-preprint server was a way of automating an existing process.

European-integration research, by contrast, even built quality-control mechanisms into its working-paper series. It may be that as the field matures, it will become more comfortable with sharing papers that have not been reviewed. Even now in European-integration research, the demand for easy access to a variety of papers quoted in bibliographies is growing. One type of grey literature, conference papers, is widely circulated via references but very difficult to obtain by those who did not participate in the conference. Those conference papers are cited even though papers submitted to conferences seldom undergo any quality check. [5] [#note5]

It seemed to us as we planned ERPA that an archive has to both acknowledge the demand to make literature of all types retrievable, and yet find a quality filter to keep readers from spending time on relatively low-utility papers.

Possible Scenarios

The possible scenarios for offering papers through an archive can be summed up in a table that integrates the degree of quality control with the types of papers included.

Ideal Types of Electronic Papers Archives	Quality Control	Types of Papers
Exclusive Club Scenario 1	Strict	Working papers only
Central Access Scenario 2	Moderate	Working papers only
Comprehensive Search Tool Scenario 3	Moderate to strict	Working papers and e-journals
Depository Scenario 4	None	All
Conference Archives Scenario 5	None to moderate	Conference papers only
E-print Server Scenario 6	None or indirect	Working papers only

I will discuss the alternative scenarios in turn.

Scenario 1: The exclusive club

The archive contains working-paper series from organizations that meet strict requirements for their review process.

This scenario is intended to keep out all but the highest-quality working-papers series by limiting the organizations whose papers are accepted to those that follow one or several of the following standards: having double-blind outside peer review, professional editing, or only top academics on the Editorial Board.

The rationale for this approach is twofold: First, the archive does not confuse readers by combining papers that have undergone a strict quality-control process with papers that have not. This archive's mission is to facilitate access to papers that are considered by the archive managers to be worth reading, and to exclude lower-quality papers. Second, the archive is attempting to raise the quality of published papers by focusing on a few participating high-quality series. The assumption is that working papers that are not easy to find (because they are not in the archive) will be less likely to be read and considered for publication in top journals, and that authors who want to be published in those journals will go through the review process of the participating institutions to get the archive's imprimatur. The fact that the exclusive club's members are top institutions and top series explains the attractiveness of the archive to both authors and readers.

> "Publishers' online offerings might be searchable, but usually only across the publishers' own journals — a criterion that is almost useless to scholars"

The problem with this scenario is that it does not help readers overcome the growing complexity of the Web. Limiting an archive to a small segment of the e-publications available in a field seems too restrictive. Furthermore, an archive that is too exclusive will not be widely used, which may frustrate the goal of broad readership.

Scenario 2: The central access point to all quality-controlled online series

The conditions for inclusion in the archive rule out unreviewed series. The aim is to make the archive the central access point to all quality online series in the field.

This second scenario is less restrictive than the first, in that the archive accepts all the papers from any quality-controlled online series. The challenge is to set the standards so that all series that are of sufficient quality are included and — by extension — recommended by the archive. Such standards might include:

- The sponsoring institution for the series is one of the best in the field based on its publication record, the key events it organizes in the discipline, or other similar criteria;
- Members of the institution are well known worldwide through their publications; or
- The series' referees are either top-ranking themselves or there is at least a serious double-blind peer refereeing process involving at least two reviewers.

Scenario 3: The comprehensive search tool for papers in a discipline or specialty

This scenario includes both working papers and e-journals in the archive.

In addition to offering all quality working-paper series, this archive also offers papers in electronic journals or the electronic versions of print journals with whom the archive operators have worked out arrangements for participation. That makes it a truly comprehensive search tool for quality papers in a field. That comprehensiveness makes participation attractive to journal publishers, whose papers become part of the wide range of online literature, accessible with the same search tool used for the rest of the literature in the discipline. That is a boon to their authors, too, who want their work to be read and cited. (Publishers' online offerings might be searchable, but usually only across the publishers' own journals, which may cover a wide range of different disciplines — a criterion that is almost useless to scholars.) Since access to the full text of journal articles is normally only granted when a subscription fee, a pay-per-view fee, or a site fee has been paid in advance, the link from the archive's search-result page might not lead to the full text or a download page, but to the home page of the journal, where there are instructions on how to access the online version of the article. As this scenario has not yet been tried in European-integration research, we don't know whether publishers would indeed be interested.

Some archives in this scenario might include selected papers from online series or journals outside the discipline or specialty that may be of interest to the community.

Scenario 4: A depository of papers in a discipline or specialty

This scenario opens the archive to any series or paper. Managers of such an archive might seek out series. The archive would become a full depository of papers in the field.

This archive is the most inclusive. Its scope extends to all kinds of papers, regardless of whether they are published in online working-papers series, in e-journals, on conference Web sites, or on private home pages; whether they represent genuine scientific research or "policy papers" (policy recommendations or explorative essays); or whether they come from a site that includes only one, a few, or an entire series of papers. The operators of this archive actively seek to include as many online papers as possible by inviting all potential candidates to participate.

This vision of an all-encompassing archive in a discipline represents the far end of a continuum that ranges from exclusive with strict quality control to inclusive with little or no quality control. Will an archive that includes papers of very different quality and types serve the community? I believe it would be a doubtful service unless the archive takes steps to overcome the Web's natural blurring of the distinctions between sources. By highlighting those distinctions, we could enhance the quality of the search. In my own field there is an example of an analogous approach: When the European Economic Community opened the cross-border market for foodstuffs in the 1980s, it only minimally

harmonized the quality-control laws and regulations in each country. Instead, it mandated strict labeling laws informing the customers about the products' ingredients. An archive could take a similar approach, giving users some clues about the type and quality of the papers they are retrieving via the archive's search engine. That could take two forms:

1. The archive's search form could include a series of options to allow the user to select the type and quality of papers desired: heavily reviewed, working papers, journal articles, etc. A user could choose one or many categories. The archive that wanted to emphasize quality could use as its default a search for papers that have undergone some sort of quality check.

2. The search-result pages could label each paper according to its type and the quality checks it has passed. The archive might develop a series of icons to attach to each paper listed, or to present the search results ordered from strict quality control to no control. Labels could include:
   Type of paper
   - in-house working paper, where the authors are members of the organization that does the selection
   - open working-paper series, where the authors are not necessarily members of the organization
   - research papers that are not part of a series
   - electronic journals
   - conference papers
   - "gray" papers such as policy reports and student papers
   Quality-control mechanism
   - none
   - ratings
   - internal refereeing
   - external refereeing
   - open peer commentary
   - external double-blind refereeing
   Publisher affiliation
   - scholarly association
   - commercial publisher
   - extra-university research institute
   - university department
   - none

The labels would be applied ex ante; that is, the series would be rated or labeled, not the individual paper. This formalized labeling mechanism would not assess the contents of a paper. Therefore categories like "research paper" vs. "policy paper" or "empirical paper" vs. "theoretically oriented paper" would not make sense, since they would presuppose reviewing the paper itself.

As already mentioned, to become truly comprehensive the archive might also include a mechanism to include preprint or off-print papers from researchers' personal Web sites. The same type of interface that would allow adding individual journal papers to the database might easily implement that. Inclusion should not be completely automatic, however: In order to avoid misuse, the archive operators should filter the papers before posting them.

Whether archive operators opt for this fourth scenario will depend largely on the answer given to the central question raised above, namely whether it makes sense in their research community to provide access to all types of papers regardless of quality checks.

Scenario 5: The conference-papers archive

This archive would offer conference organizers the opportunity to register and eventually archive papers presented at their conferences.

Conference papers tend to be more ephemeral than other scholarly work. Even if they are available electronically, they often are spread over numerous sites that are themselves not always maintained for long periods. Finding them often means tracking them down one by one. A conference-papers archive would establish a one-stop access point. The operators of the archive could propose to the conference organizers two solutions: either the organizers upload all papers to the archive — sometimes even making the archive version the "official" one; or the archive organizers might use "spider" technology to extend the archive search engine to the conference site, and allow it to retrieve papers from that site. (Such a search would typically retrieve papers by title, conference venue, dates, and authors; more sophisticated searching would require adding metadata markup to the papers and a far more elaborate search interface, which could be expensive.) See the sidebar to this story to find out how ERPA handles the technical details.

Scenario 6: The dynamic e-print server

Anyone could make a paper available to the community under this scenario; there would be no review mechanism beyond the registration. The service is dynamic in that papers can be replaced or updated as new versions are created.

"A basic test is whether it is possible to attribute a number of ERPA keywords to

This scenario is similar to Paul Ginsparg's physics and mathematics archives in Los Alamos. Ginsparg says that his model has a level of quality control: "The archives do benefit from an automatic form of peer review, since users typically replace their submissions in response to direct feedback, and subsequent revisions frequently benefit as much or more from this feedback

the papers"   as from the conventional referee process" (Ginsparg 1998
[https://quod.lib.umich.edu/j/jep/3336451.0005.109/--european-research-papers-
archive-quality-filters?trgt=p_ginsparg;view=fulltext#p_ginsparg]). Eventually, he
says, a "global raw research archive" could grow out of his model, and he suggests that there could eventually be "a variety of superficial improvements" that would have the kind of labeling on each paper that is similar to the labels in Scenario 4 (but the labels would be given on the basis of the individual paper, not of an ex ante assessment of the source of the paper). "Any type of information could be overlaid on this raw archive and maintained by any third parties," he writes. He thinks labels might also indicate recommendations ("essential reads" for a given subject); grading according to overall importance; quality of the research; information on follow-up research, or even the successful (or unsuccessful) submission to a formal refereeing process. Those labels could be assigned by anyone in the community (but it would require a sophisticated system that balances widespread participation against avoiding bias).

ERPA's Policy Paper

ERPA was originally conceived as an exclusive club, presenting only series from its members. However, in assessing the alternatives we realized that we would serve the community better if we opened up that exclusive club. We decided, instead, to focus on a smaller set of quality series to build a "centre of excellence" that would help make European-integration studies more accessible. We also decided not to try to include e-journals, since that would mix commercial and non-commercial activities. We acknowledged the demand for easy access to conference papers, but considered this a project outside the scope of the participating institutions. (Including conference papers would also mean that the ERPA software would have to be reworked considerably. because it was not designed to give access to more than a few thousand of papers.) All our work is detailed in the policy paper [http://eiop.or.at/erpa/policy.htm] issued in June 1999.

ERPA settled in the middle ground, becoming a central access point to all quality-controlled online series (Scenario 2). Series have to meet the following criteria to be eligible for inclusion in ERPA:

- The contents of the papers have to relate to European-integration research in the wider sense. A basic test is whether it is possible to attribute a number of ERPA keywords to the papers. If a series includes both European-integration papers and others, only the European integration ones may be included in the Archive.
- The series must have published at least three papers on line per year for the last two years. (Those numbers apply to the European-integration papers and not to the overall number of papers in a series.) By "on-line publication" we mean full-text publication in HTML, Postscript or PDF. Paper publication with only tables of content or abstracts on line are not eligible.
- We do not encourage (but we do accept) applications from series whose papers are not in one of the major, widely spoken European languages. In any case, the metadata (descriptions, keywords, etc.) has to be in English.
- On the basis of an overall assessment, the papers should conform to scientific standards, e.g. scientific style, footnotes, and bibliography. It does not matter if a minority of the papers in the series falls in the category of "policy papers" — but pure policy-paper series are not eligible.
- All papers of participating series have to undergo a quality check that involves more than one person through an internal or external refereeing system. The series publisher is asked to describe the reviewing system in detail, and that description is posted on the ERPA site.
- The series publisher must be a university unit, an extra-university research institute, or a scholarly association. Series edited by individuals or groups of individuals are not included in ERPA.

The management board reviews all applications and makes its decision based on the policies detailed in the ERPA policy paper [http://eiop.or.at/erpa/policy.htm] .

Outlook: Quality Control in the Digital Age

Until the advent of the Internet, with its opportunities to publish with much less effort and at much lower cost than in printed formats, the market for academic publications was quite different. Working papers played a minor role in academic discourse because distributing them was so difficult. Journals and books needed commercial publishers and, therefore, outlets for scholarly work were more limited. While publication in print through a publishing house was by no means a guarantee of high quality, the double bottleneck of a limited number of publishers and of a fixed number of papers per journal issue created an environment of competitiveness. That usually meant higher quality, but it also meant slower publishing, since the top journals always had the most thorough review.

Now electronic publications are widespread in a growing number of disciplines and the situation has changed. An increasing number of sites offer academic papers without submitting them to a rigorous filtering process. These working papers, once available to only a small group, are now easily retrieved worldwide. Most important, the models are a first step. The door is open for even more imaginative solutions, combining both more transparency and comprehensiveness with more freedom of choice.

Dr. Michael Nentwich is a senior researcher at the Institute of Technology Assessment [http://www.oeaw.ac.at/ita/] of the

Austrian Academy of Sciences in Vienna where he is mainly involved in projects in the area of information and communication technologies. He is also involved in a number of WWW projects and edits the European Integration online Papers (EIoP) [http://eiop.or.at/eiop/] . Previously, he was a lecturer at the interdisciplinary Research Institute for European Affairs [http://fgr.wu-wien.ac.at/institut/ef/ief-home.htm] at the University of Economics in Vienna, an HCM fellow at the Universities of Warwick and Essex in the U.K.. He is currently a guest researcher at the Max Planck Institute for the Study of Societies [http://www.mpi-fg-koeln.mpg.de/] in Cologne, Germany. He studied law, economics and political science in Vienna and Bruges/Belgium. His publications include books and articles on European economic law, European constitutional issues, democratic theory and technology assessment.

Dr. Nentwich's home page is at: http://fgr.wu-wien.ac.at/nentwich/mn.htm [http://fgr.wu-wien.ac.at/nentwich/mn.htm]

Michael Nentwich may be reached by e-mail at mnent@oeaw.ac.at [mailto:mnent@oeaw.ac.at] .

Footnotes:

1. It is not the place here to expand on the topic of growing e-publications; but see, for instance, the ARL Directory of Electronic Journals, Newsletters and Academic Discussion Lists, [formerly http://db.arl.org /dsej/index.html] [#note1-ptr1]

2. See, for instance, the steadily growing collection of URLs at the EuroInternet Web site, [formerly http://fgr.wu-wien.ac.at/nentwich/euroint1.htm - wp] [#note2-ptr1]

3. E.g., the Mannheim ECPR Joint Sessions of Workshops 1999, [formerly http://www.essex.ac.uk /ECPR/jointsessions/manprog.htm] [#note3-ptr1]

4. In comparison, the widely known European Integration online Papers, edited under the auspices of the European Community Studies Association Austria, has more than nine hundred subscribers worldwide in their third year. (EIoP subscription is free). [#note4-ptr1]

5. In general the abstracts are filtered, but the papers themselves get no review until the discussion at the conference. Those discussions have no immediate impact on the papers that are included in a conference Web site. [#note5-ptr1]

References:

Ginsparg, Paul. 1998. "Electronic research archives for physics." In: I. Butterworth (ed.). The impact of electronic publishing on the academic community, an international workshop organized by the Academia Europaea and the Werner-Gren Foundation. London/Miami: Portland Press. http://tiepac.portlandpress.co.uk/books/online/tiepac /session1/ch7.htm [http://tiepac.portlandpress.co.uk/books/online/tiepac/session1/ch7.htm]

Nentwich, Michael. May 1999. "Cyberscience: Die Zukunft der Wissenschaft im Zeitalter der Informations- und Kommunikationstechnologien." Working paper of the Max Planck Institute for the Study of Societies no. 99/6. Available as a PDF file [http://www.oeaw.ac.at/ita/ebene5/wp-mpifg.pdf] .)

Links from this article:

Academy of European Law and the Robert Schuman Centre of the European University Institute in Florence, Italy, http://www.iue.it/ [http://www.iue.it/]

Advanced Research on the Europeanisation of the Nation-State (ARENA) http://www.arena.uio.no/ [http://www.arena.uio.no/]

European Research Papers Archive (ERPA), http://eiop.or.at/erpa/ [http://eiop.or.at/erpa/]

EuroInternet Web site, [fprmerly http://fgr.wu-wien.ac.at/nentwich/euroint1.htm]

European Integration online Papers (EIoP), based in Vienna, Austria, http://eiop.or.at/eiop/ [http://eiop.or.at/eiop/]

Harvard Jean Monnet Chair in Boston, http://www.jeanmonnetprogram.org/ [http://www.jeanmonnetprogram.org/]

Max Planck Institute for the Study of Societies in Cologne, Germany, http://www.mpi-fg-koeln.mpg.de/ [http://www.mpi-fg-koeln.mpg.de/]