

Aeroflot proposal

Thomas Krichel

2001–11–23

1 Introduction

0. This document is the Aeroflot protocol. Its initial version was written on flight SU315 between Moscow and New York on 2001–09–19 by Thomas Krichel, for submission to the NEP editor list. The latest version may be found at <http://openlib.org/home/krichel/aeroflot.html>. It has benefitted from comments by Bernado Batiz-Lazo, Mark Fitzgibbon, Bernd Hayo, Sune Karlsson, Victor M. Lyapunov, Marco Novarese, Sergei I. Parinov, Murat Yildizoglu and Christian Zimmermann.

1. NEP is a RePEc user service that filters new additions of working papers into subject specific reports series. The filtering is carried out by a group of volunteer editors. The reports are circulated by electronic mail. The internal operation of the service is governed by the York protocol.

2. NEP was founded by John S. Irons and Thomas Krichel in 1999. The technical support infrastructure has been managed by José Manuel Barrueco Cruz. The mailing lists are hosted by JISCmail. In June 2001, there were over 40 lists serving more than 5,500 individual email addresses. NEP is coordinated by Bernado Batiz-Lazo, who holds the title of general editor at the time of writing.

3. The RuPEc team, coordinated by Sergei I. Parinov at the Siberian Branch of the Russian Academy of Sciences, is a grantee of the Ford Foundation for the Socionet project. This project has build a RePEc user service that emphasizes the timeliness of RePEc documents as well as customization and filtering tools.

4. It is obvious that the two efforts can benefit from each other. They both attempt to solve the same problem of filtering the stream of documents into subject-specific or user-specific sets. The NEP project is about humans performing the filtering. Socionet emphasizes computer-based filtering. In order to save human labor, it is a good idea to investigate if and how computers-based methods can help human-based methods.

5. It is proposed that RuPEc build a new support infrastructure for NEP. This system is initially code-named “nepsi”.

6. The most important design principle for nepsi is that the community of NEP editors be involved as much as possible in the design and running of the service.

7. Nepsi will be produced in two phases. In phase 1, the traditional NEP system will be maintained in parallel to the nepsi system. If the NEP editors agree, in phase two nepsi will become the single interface for NEP report editing.

8. This document only contains design goals and implementation details for phase 1, in Sections 2 and 3, respectively. For phase two, there are only some indicative ideas in Section 4. Technical details are confined to Section(sec:tech).

2 Design goals for phase one

9. Nepsi will be a web site for the use of NEP editors. Editors will use the site to go to inspect data about new documents that may be included in a report

10. Nepsi will free editors from the weekly routine of editing list. Editors will be free to issue reports as and when they see find.

11. Nepsi will aim to work towards making the size of reports more equal. Instead of issuing a draft report every week, nepsi will send an email alert to the editor when the number of papers that have not yet been scrutinized reaches a critical level. The value of the critical level is decided by the editor and can be set at the web site.

12. Nepsi will check that the full-text URL of incoming files is valid. Validity means that a full-text document can be freely withdrawn or that the document could be withdrawn under restrictions. Nepsi will aim to distinguish between these cases. New papers arriving in RePEc will not be made available to editors until such time as a valid full-text URL has been found. This is to avoid the presence within NEP reports, of dead links to full-text of papers.

13. Nepsi will attempt to make editors' work a little easier by making a computerized guess of which papers are most likely to be included in the report, and sorting them accordingly.

14. Nepsi will keep a backtrack of all previous selections of the editors. If the editor agrees, these will be exported to RePEc user services and may be helpful to them.

15. NEP editors will be given the opportunity to use nepsi facilities to rank articles. Top-ranked article will be arriving at the top of the report. The value of the rank of each paper may be included in the report.

3 Implementation details for phase one

16. NEP editors use the HoPEc service to announce their personal details to nepsi. In addition, the general editor of NEP will have a special login that will allow them to associate NEP editors to their lists.

17. At the start of the service, RuPEc will collect historical information about all previous editions of NEP reports for each list, to the extent that such information is available from JISCmail or other sources. Throughout the running of the service, nepsi will keep a trace of all the NEP report issued.

18. When the NEP editors enter the nepsi site, they will find all the new additions that have occurred since the last report. These additions are sorted by the interface to start with the one that is most likely to be found in the next edition of the report. The aim is to build a system that learns with each issue of the report what new items are most likely to appear on the report. How this is to be done will be a major research issue of the project.

19. The report issues will be disseminated as text email. The email will be sent to the address of the NEP editor for forwarding to the mailing list.

20. NEP editors will need to enter the nepsi site at periodic intervals agreed with the NEP general editor. These intervals will be set to give flexibility in the generation of reports rather than imposing fixed periodic times for dissemination of new additions. When no reports are issued, the number of new papers simply accumulates in the account of the NEP editor.

21. Editors will be reminded by email agreed time for an issue is coming up. If the number of incoming, not disseminated papers reaches a critical level, both the editor concerned and the general editor will be warned.

22. The general editors will receive extensive reports from nepsi to check that the editors are doing a proper job.

4 Design for phase two

23. If the editors agree and a suitable technical solution can be found, the mail list management will be taken over by nepsi. Nepsi will then be the single interface of interaction between editors and NEP. In addition the nepsi system as envisaged for phase two may incorporate one or more of the following features.

24. Nepsi of phase two may become an integrated tool for both editing of reports and management of list membership. It is expected that the Mailman software will be used to allow for list management using nepsi. Mailman allows for a complete list management through a web interface. Both the NEP editors and NEP users will find such a system easier to work with than the JISCMAIL-based system.

25. Nepsi of phase two may use redirection scripts to log downloading activity coming from report data. This would enable nepsi to contribute download data to LogEc.

5 Technical details *subject to revision*

26. Reports will use the AMF representation of RePEc available through an the OAI gateway at <http://oai.repec.openlib.org>. It will used ReDIF data as converted to AMF, using UTF-8 encoding.

27. The back issue of a report should contain the full ReDIF metadata valid at the time the report was issued, using the AMF `ref=id` construct to mark this data as non-authoritative.

28. Reports will be represented in AMF through a two-level nested collection noun of AMF. One collection noun is used per NEP issue. An other collection noun is used for the report that collects all issues. This metadata will be refereed to as report data.

29. Report data will be mirrored to the OAI gateway and be made available through it.