

Report on content acquisition, dissemination and service specifications



**ECP-2006-DILI-510001**

**NEEO**

**Report on content acquisition,  
dissemination and service  
specifications**

<b>Deliverable number</b>	<i>D2.3</i>
<b>Dissemination level</b>	<i>Public</i>
<b>Delivery date</b>	<i>28 March 2008</i>
<b>Status</b>	<i>Final</i>
<b>Author(s)</b>	<i>Carlos Torres-Vitolas, Michelle Blake, Frances Shipsey, Michael McFarlane, Tim Green</i>



**eContentplus**

This project is funded under the eContentplus programme<sup>1</sup>,  
a multiannual Community programme to make digital content in Europe more accessible,  
usable and exploitable.

---

<sup>1</sup> OJ L 79, 24.3.2005, p. 1.



## I. Introduction

This report presents the needs and demands of potential end users of Economists Online (EO) regarding the access and use of online research material in Economics identified by the User Requirement Report (Torres et al., 2008). In addition, it highlights the suitability of the current work plan to satisfy such demands as well as the possible refinements necessary to achieve such an objective.

The specific issues tackled in the present document were identified from a consultation of researchers, academics and students from the NEEO partner institutions and economists working in the public and private sectors between September 2007 and January 2008. More specifically, the User Requirement Report, via quantitative (an online survey applied to 533 respondents) and qualitative methods (3 focus groups and 1 blog), allowed the identification of current Economics researchers' needs for access to full text documents and statistical datasets; their needs for storing and disseminating their research material; the types of services they consider important for implementation in EO as well as relevant multilingual and multicultural issues that may condition current researchers' practices.

Based on the evidence obtained, this report specifies the aspects the service provider will need to consider addressing either during the project or after, as part of the Business and Sustainability Plan, so as to meet the needs and expectations of potential users in a satisfactory manner. For this purpose, it provides confirmation of the programmed activities in the established NEEO work plan - Grant Agreement Annex 1: "Description of Work" - and further recommendations on the content to be aggregated from Economics researchers, how Economics researchers would like to see it disseminated and relevant service specifications.

## II. Objectives

To provide a written plan for the acquisition and dissemination of content as well as for service specifications derived from the User Requirement Report, which includes recommendations on the following aspects:

- Content to be aggregated from EO users;
- Prospective routes and practices for content dissemination;

- Set of service specifications

## III. Recommendations

### 3.1. Content

#### 3.1.1. Content: traditional publications.

The current work plan establishes the target of providing 50,000 bibliographic references and links to 15,000 full text documents by the end of the project (Annex 1 – Description of Work). The types of publications to be disseminated by EO are intended to be very comprehensive and to respond to economists’ research and teaching needs by including articles, working papers, books, book chapters, theses and conference papers.

The evidence obtained from the User Requirement Report showed that there are different needs of access according to the type of publication and, furthermore, that different publication types do not have equal value for economists. More specifically, respondents to the online survey appeared to have good access to journal articles while books, book chapters and datasets were considered the most difficult types of research material to access online accompanied, to a lesser degree, by theses and conference proceedings. However, focus group participants expressed different levels of interest in accessing these scarce types of research material in the following order: books and book chapters, datasets as well as non-peer-reviewed publications (e.g. conference proceedings and working papers).

Consequently, it is important that, as part of the current work plan, EO aims at filling in the gap in the provision of online open access to books and book chapters (very limited or non-existent among online academic repositories such as Research Papers in Economics - RePEc - and the Social Science Research Network - SSRN). For this purpose, it is recommended NEEO partners encourage contributors from their institutions to submit permitted open access versions of those to their respective institutional repositories.

Focus group participants indicated that non-peer-reviewed material such as conference papers and proceedings, as well as working papers, play an important role in providing up-to-date information about ongoing research projects



conducted by authors and institutions. They allow researchers to follow the development of methodologies and analyses by key fellow economists. Common limitations identified which restrict their availability online include the fact that conference proceedings are not usually stored online for long periods of time by conference organisers. Usually only summaries or abstracts are provided rather than full-text documents. Furthermore, although academic working papers do not appear to be very difficult to access online, consultancy reports conducted for firms and governmental institutions are not easily accessible.

In order to address these circumstances, it is recommended that the leading economists identified as part of Work Package 3 are invited to submit recent conference papers or proceedings they have produced that they consider reflect their current on-going research projects. If such material is openly accessible online through EO, it would help to provide a more up-to-date and comprehensive representation of NEEO partners' work.

As part of the work plan for Work Package 3, it would be advisable to incorporate within EO conference materials produced for recent Economics events organised by NEEO partner institutions. In addition it is also recommended that, as a general policy, NEEO partners include full-text conference proceedings and papers in their repositories as a priority in preference to summaries or abstracts where possible.

It is also advisable to encourage authors whose publication lists will be included in EO to submit consultancy reports they consider to have academic relevance for fellow researchers while not violating any confidentiality agreements with their corresponding clients. These reports would serve not only as sources of information to fellow researchers on key case studies but also as illustrations of the practical applications and policy implications of the academic research conducted by NEEO researchers.

The User Requirement Report identified that potential end users of EO were much more interested in accessing the published version of a research output than the submitted version. Participants feared that the latter may differ from the final published version and, hence, may constitute unreliable sources of information for reference purposes. This need has already been addressed by the project and, as stated in the First report on IPR issues (Reid, 2008); consequently, it has been stated that the NEEO project will target primarily refereed accepted versions of



publications. However, it is recommended that, when technically possible, end users are provided with links to the fee-based online services that store the published version of the publication for both pre-refereed and refereed author versions of research outputs. In this manner, economics researchers entitled to access the latter would be able to refer and work with the ideal version of the publication.

In addition, the current work plan also stated that the central gateway will aggregate metadata of Economics publications from other sources than the institutional repositories of the NEEO partners. These include RePEc alongside a selection of author / institutional web sites of economic content. In that respect, the evidence obtained from the User Requirement Report identified SSRN as a commonly used online resource for accessing research material on Economics; it would be important, therefore, to consider the latter as a potential source of bibliographic references if the information is available.

Finally, the information obtained from potential end users of EO has also shown the significance of periodically verifying the active status of links contained in the publication lists of the service. Although NEEO partner institutions will be in charge of maintaining their respective institutional repositories, after the service starts aggregating metadata from sources other than NEEO partners, this could potentially constitute a continuous and ever growing task. Therefore, it is recommended that this issue be assessed in the future Business and Sustainability plan.

### **3.1.2. Content: Datasets**

According to the current work plan, one of the project aims is the aggregation of metadata of NEEO partner dataset; with each of the partners incorporating at least 10 datasets (i.e. a total of 160 datasets). In principle, the project will include only original datasets developed by researchers working in the NEEO partner institutions, allowing open access to end users while avoiding copyright infringements. In addition, it is planned that this information will be aggregated alongside the respective publications based on such data so that end users will be able to navigate from one type of document to the other and vice versa.

The information presented in the User Requirement Report confirmed the great interest among Economics researchers in accessing datasets. Focus group



participants indicated that such an input would allow the service to position itself rapidly among the most popular online resources in the field. However, the needs of economists' access to datasets are clearly vast and are likely to surpass the planned objectives outlined in the current EO work plan. This is because Economics researchers are not only interested in accessing datasets developed for academic purposes but also those of a commercial nature. Furthermore, they would like to see a portal of data and software extensions for technical analyses implemented as a feature of EO.

Consequently, it is recommended that the project considers, in its Business and Sustainability plan, the development of a dataset portal which, alongside the datasets previously aggregated into EO by NEEO partners, may offer links to other data collections containing international (e.g. International Monetary Fund, The World Bank or the European Union data portals), national statistics from academic and governmental institutions (e.g. national institutes of statistics) as well as to other online resources for searching datasets (e.g. the National Bureau for Economics Research - NBER - or the World Wide Web Resources in Economics – WEBEC). The development of a worldwide index of datasets for Economics after the end of the NEEO project is expected to contribute towards meeting this objective.

As mentioned above, the evidence obtained from potential end users revealed that the data interests of economists are not limited to datasets but also include associated tools such as command or syntax files as well as technical software extensions and/or applications developed by researchers with innovative methodologies for data analysis. It is recommended to incorporate in the current work plan provision of these technical components alongside the associated datasets on which they were used. In this manner, end users of EO will be able to replicate original results and have a deeper understanding of the research's methodology.

However, the capacity of the service provider to aggregate datasets from contributors to EO on a regular basis faces some challenges due to some concerns among dataset creators. Although only 3% of survey respondents said they were not willing to provide any of their datasets, most of them wanted certain conditions in order to collaborate effectively with the project such as a clear statement of creator(s) rights (75%) and a clear statement of end-user(s) permissions (65%). As identified by focus group participants, datasets creators have two key worries: fears

of not being adequately acknowledged by EO end users; and fear of losing publication opportunities if other researchers, with more resources, decide to work on similar analyses simultaneously.

In that respect, the intended linkage between the metadata of publications and the metadata of datasets included in the work plan would constitute an important feature of the service. The possibility of retrieving full references and, possibly, the full-text of the study produced with this data would promote the correct acknowledgement of dataset developers.

An element not contemplated in the current work plan that could serve to reassure dataset creators to submit their data is the establishment of a registration procedure at the moment of downloading datasets which would include the commitment to respect a license agreement. In addition, it is recommended that an automated email containing a copy of the license agreement is provided to users at the moment of registration. This should be done where the data is being stored by the local repository and is not a NEEO responsibility. This license would need to state clear commitments for end users, including not using such data for commercial purposes without the consent of developers, to fully acknowledge the dataset creators in any publication or piece of research resulting from the use of the data and not to appropriate in any other form of its intellectual property. However, NEEO partners would still have responsibility for correctly labelling and securing the datasets themselves. This could therefore be an additional safeguard for dataset creators.

### **3.1.3. Content: Copyright concerns**

As part of the established work plan for Work Package 3, it has been agreed that a copyright toolkit will be developed with information about basic legal frequently asked questions (FAQs) and containing model contracts. This will help to ensure the safe and confident delivery of content by authors into the corresponding institutional repositories for the project. In addition, it has been agreed that the toolkit will be adapted for specific national jurisdictions and distributed by partners. The Work Package 6 team will translate the resulting toolkit from English into the three alternative languages of NEEO: French, Spanish and German.

The evidence obtained from the online survey and focus groups signals that the development of the aforementioned toolkit and its adequate dissemination are of



great importance for the success of the project. Both Economics researchers and dataset creators feared they may infringe copyright laws by non compliance with their agreements with publishers (to whom Economics researchers submit not only the article but also their datasets and key additional technical files for replication purposes). In addition, economists were unsure to what extent their datasets constitute an original piece of work if they are based on re-elaborations and transformations of copyrighted protected data<sup>2</sup>.

In response to such needs, and as already stated by the current Description of Work, it is recommended to complement the local distribution efforts conducted by the NEEO institutional partners by placing links to download the toolkit within the EO website in its English and translated versions (French, Spanish and German). In addition, as part of the work of WP5 the copyright knowledge bank of SHERPA ROMEO<sup>3</sup> will be integrated into local institutional repositories. This will allow potential contributors to verify the policies and conditions set by their publishers in more detail.

Finally, as stated in the previous section, datasets intended for inclusion in EO in EO initially will be original products from researchers working in NEEO partner institutions. However, in consideration of the future expansion of the service it is important that there will also be a copyright toolkit addressing concerns specifically related to datasets. This action has also been agreed upon in the First Report on IPR Issues and it is expected that Work Package 4 (Content: datasets) will produce it by January 2009. In response to the concerns identified in the User Requirement Report, it will be important to address the following key aspects: creator(s)' rights over data submitted to publishers and originality of indicators and measures developed from copyrighted data.

### 3.2. Dissemination

For the purpose of disseminating the content of EO, the NEEO project has established two key lines of activities. These are: the diffusion of EO content through key Economics online service providers; and the development of an advocacy

---

<sup>2</sup> The Directive 96/9/EC of the European Parliament and of the Council of 11 March 1996 on the legal protection of databases should also be mentioned. This is a European Union directive in the field of copyright law which harmonizes the treatment of databases under copyright law, and creates a new sui generis right for the creators of databases which do not qualify for copyright. More information can be found at:

<http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:31996L0009:EN:HTML>

<sup>3</sup> <http://www.sherpa.ac.uk/romeo.php>





programme that will advertise EO content widely among economics researchers. The latter includes the development of publicity materials, presentations and exhibitions as well as articles submitted for publication in professional journals and local institutional newsletters.

In relation to the first of these actions, the work of the Work Package 5 team (Interoperability, Infrastructure and Gateway) in creating a RePEc archive and enabling additional access through Google and Google Scholar is of great importance. The information obtained from Economics researchers through the online survey identified Google (65%), Google Scholar (70%) and RePEc (67%) as the most preferred online services where researchers want to see their work disseminated. It is recommended, in addition, to also consider the Social Science Research Network (SSRN) as a possible service to target (57% of survey respondents expressed being interested in seeing their work distributed through this online service), if technically possible.

In the User Requirement Report it was noted that known authors' websites are also commonly targeted by economists to access research material online (57% of survey respondents). It is advisable, therefore, that the leading researchers working in the NEEO partner institutions, identified as part of Work Package 3, are encouraged to use EO's publication lists and the corresponding links in their own personal websites.

It is recommended that the advocacy programme developed to advertise EO content, when participating in presentations and exhibitions or placing articles in newsletters, emphasises different features of the service in response to local audience needs. For this purpose, it is advisable that the content of such materials follow the particular interests and needs identified across the European countries by the User Requirement Report. For example, it would be useful to stress the great need of accessing datasets manifested by respondents from Belgium, Czech Republic and The Netherlands when addressing such an audience. Likewise, it would be useful to address the great interest manifested by respondents from France and Germany in disseminating their work through SSRN.

Finally, it is recommended that the advocacy programme publicising EO emphasises the added value that it provides to European Economics researchers in comparison to similar online service providers. This would help to position EO rapidly

among researchers and to differentiate it from other online repositories. In consideration of the needs expressed by Economics researchers, it would be important to highlight the following features: the greater visibility it provides to European researchers and institutions compared to American centred sites, the improved visibility of their work in searches through Google, Google Scholar and/or RePEc, access to datasets, its multilingual functionality, constant updates of operative links, access to free full texts as well as any other services that are incorporated into EO and are not currently offered by similar initiatives.

**Table 1. Summary of recommendations and status in the work plan**

The below table summarises the recommendations from the Report on content acquisition, dissemination and service specifications. There are three statuses and the status of each recommendation has been indicated in the status column. These are:

- *Currently in the work plan*: this indicates that the recommendation has already been envisaged as part of the current work plan.
- *Yet to consider in the current work plan*: this indicates that the recommendation will be discussed by the NEEO partners at the Project meeting in month 10 (June 2008).
- *Yet to consider in the Business & Sustainability plan*: this indicates that the recommendation is out of the scope of the NEEO project but will be considered as part of the Business and Sustainability plan. The first version of this is due in month 15.

Recommendations	Status
<b>3.1.1 Content: Traditional publications</b>	
a. EO offers the following publications which are currently scarce online: books, book chapters, conference proceedings.	<i>Currently in the work plan.</i>
b. EO encourages authors to provide open access versions of books and book chapters where possible.	<i>Currently in the work plan.</i>
c. EO targets author versions of publications.	<i>Currently in the work plan.</i>
d. EO aggregates metadata from RePEc	<i>Currently in the work plan.</i>
e. Leading economists identified by EO are asked to submit recent full text conference papers or proceedings that reflect their key current research projects.	<i>Yet to consider in the current work plan.</i>
f. EO encourages contributors to provide consultancy reports of academic relevance.	<i>Yet to consider in the current work plan.</i>
g. EO provides links to fee based online services that store the published version of a paper alongside links to pre-refereed and refereed author versions if technically possible.	<i>Yet to consider in the current work plan.</i>



Recommendations	Status
h. EO includes conference proceedings organised by NEEO partners.	<i>Yet to consider in the Business &amp; Sust. plan.</i>
i. EO aggregates metadata from SSRN, if information is accessible.	<i>Yet to consider in the Business &amp; Sust. plan.</i>
j. EO continuously verifies that links to full-texts from other sources than NEEO partners are active.	<i>Yet to consider in the Business &amp; Sust. plan.</i>
<b>3.1.2 Content: Datasets</b>	
a. EO offers a collection of downloadable datasets.	<i>Currently in the work plan.</i>
b. EO provides metadata of datasets.	<i>Currently in the work plan</i>
c. EO links publications to datasets.	<i>Currently in the work plan</i>
d. EO aims to provide links to syntax or command files and software applications developed by researchers.	<i>Yet to consider in the current work plan.</i>
e. End users registered in EO receive an automated email with a copy of the license agreement.	<i>Yet to consider in the current work plan.</i>
f. EO will provide, after the project finishes, links to datasets from other institutional collections and web resources.	<i>Yet to consider in the Business &amp; Sust. plan.</i>
g. EO develops a portal of datasets and software extensions for economists.	<i>Yet to consider in the Business &amp; Sust. plan.</i>
<b>3.1.3 Content: Copyright concerns</b>	
a. EO develops and disseminates a copyright toolkit for potential users.	<i>Currently in the work plan.</i>
b. The EO website contains links to a copyright toolkit in English, French, German and Spanish.	<i>Currently in the work plan.</i>
c. SHERPA RoMEO knowledge bank will be integrated into local repositories, providing information on publishers' conditions and policies.	<i>Currently in the work plan.</i>
d. NEEO partners provide copyright information on datasets issues as part of the existing NEEO tool-kit.	<i>Currently in the work plan.</i>
<b>3.2 Dissemination</b>	
a. EO enables further visibility of European economists' work via RePEc, Google and Google Scholar.	<i>Currently in the work plan.</i>
e. The advocacy programme will emphasise the added value of EO services in comparison to existing online repositories in Economics.	<i>Currently in the work plan.</i>
c. EO encourages contributors from NEEO partners to use EO publication lists and links in their personal websites.	<i>Currently in the work plan.</i>
d. The advocacy program that advertises EO emphasises different features of the service at dissemination events in response to local audience needs.	<i>Currently in the work plan.</i>
b. EO is disseminated through SSRN, if technically possible.	<i>Yet to consider in the current work plan.</i>



## IV. Services

In the Description of Work and 1<sup>st</sup> Report on the Service Oriented Architecture (SOA) design (Place, 2008) it has been established that the core of the Economists Online (EO) portal will consist of a metadata and full text search service with a multilingual interface and additional services such as publication lists and RSS feeds. More specifically, EO's infrastructure is planned to consist of a central gateway based on existing technologies and open standards such as the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) for harvesting, search/retrieve via URL (SRU) for online search queries and a crawler that returns object files from repositories. This section looks at the requirements of potential end users of EO and the specifications defined by the current NEEO Technical Guidelines (Pauwels, 2008) and Architecture Design. It also highlights where user requirements are not covered by the current work plan.

### 4.1. Metadata ingest

The current work plan states that the central gateway will host a searchable database feed through regular harvesting of stored metadata records and object files from the Institutional Repositories (IR) of the NEEO partners. It will also harvest metadata records from a RePEc archive and non-NEEO repositories with economics content. In addition, it is planned that the resulting well-defined metadata will be enriched by automatically generating Journal of Economics Literature (JEL) classification codes and extracting reference lists. These two outputs will be added as new digital items to the gateway's metadata store.

The specifications from the work plan indicate that the aggregation of metadata solely from NEEO partner institutional repositories' is unlikely to fulfil the needs of European economists. The online survey implemented for the User Requirement Report identified that economics researchers consult sources such as RePEc (65%), known author's homepages (57%) as well as SSRN (50%) for their own research. Consequently, the planned creation of a RePEc archive by Work Package 5 will complement the metadata aggregated directly from the local IRs. However, in order to provide more comprehensible information to EO's end users, it is recommended that - if technically possible - the central gateway also harvest the content of the Social Science Research Network (SSRN).



The work plan includes activities which will help to meet the needs of potential end users. These include the described enrichment of metadata; the production of publication lists using author identifiers; and the addition of JEL codes to metadata records.

The production of quality metadata will be key for the success of the service. However, this may create a tension with economics researchers who expressed an interest in a system that requires minimal effort (in both time and technical knowledge) to submit publications manually. This means that the aggregation of detailed information could become a growing burden on local repositories and EO staff in the long term. Therefore, it is advisable that each institutional repository develops a flexible submission process, which facilitates an appropriate balance of work between researchers and administrative staff. It is recommended to consider two possible scenarios in the Business and Sustainability Plan to address this issue: i. harvesting of published works in a semi or automatic manner, or ii. linking repositories directly into institutions' content management systems so as to enable automatic capture at the time of first publication.

The User Requirement Report showed that the ingest formats compatible with the aggregation of quality metadata within EO matches closely those used by economics researchers. In the online survey, the two most heavily used file formats were MS-Word and Portable Document Format (PDF) both for the production and consultation of research material (over 80% in each circumstance). To a lesser degree, Text (txt) and LaTeX formats were also commonly used.

According to the current Technical Guidelines the following formats will be full text indexed by the service:

- PDF: application/pdf
- ODT : application/vnd.oasis.opendocument.text
- TXT: text/plain
- HTML: text/html
- [La]TeX: application/x-latex
- PostScript: application/postscript
- MS-Word: application/msword

The only file type which users indicated using which has not been specified is PowerPoint. It is therefore recommended that WP5 consider including PowerPoint as a file type to be full text indexed.



NEEO partners will be expected to describe datasets stored in their institutional repositories using the Data Documentation Initiative (DDI) standard. It is necessary to take into consideration that the most commonly used dataset formats identified in the online survey were MS-Excel, Comma Separated (CSV), Stata and SPSS. Therefore, it is advisable for local institutional repositories to support these files types so they can be harvested as part of EO.

## **4.2. Datasets**

Work Package 4 (Content – datasets) is expected to harvest the DDI metadata so as to include the datasets information into the EO portal as well as to show cross-links between publications and data. However, cross-links between publications and data will be completed at the local institutional repository level. These cross-links are important as they may help to ease dataset creators fears about not being adequately acknowledged by end users.

Due to the complexities involved with datasets and the long term possibility of fulfilling the request of potential end users of EO to develop a dataset portal it is advisable that partner institutions administer and maintain a dedicated data repository or that they have arrangements with regional or (inter)national data archives for storing data sets.

The demanded copyright safeguards by dataset creators may add to the burden local repositories and EO face in the future. A growing dataset collection would make it necessary to request and enforce relevant licenses, confidentiality agreements and limitations of access for each dataset in EO. It is recommended, therefore, that local institutional repositories think carefully about how to address this; working together with data archives or data curation centres is an option that must be seriously considered. The Business and Sustainability plan could assess the needs of partners to develop an electronic rights management system, which would allow the enforcement of licenses by controlling release and cessation dates of files as well as operate embargoes when required. In addition, it is recommended to include as part of this process the need of each local repository being able to carry out an audit of who has accessed which datasets.

## **4.3. Searching**

According to the current work plan, the enrichment of quality metadata will help to develop various search features within EO so as to suit end users demands. Work Package 5 is committed to providing publication lists based on author identifiers as well as to enhance metadata records with JEL codes and reference lists. They will also be producing relevance ranking of search results based on download rates of full-texts, allowing searches through keywords in the metadata as well as in the full text publications, grouping duplicates in search results sets and allowing end users to navigate on the metadata (e.g. find publications with the same JEL code, publications that cite the same publication or author, publications by same (co)author, etc.).

Those basic search features defined by the work plan fit most of the demands of potential end users of EO reported in the User Requirement Report. When respondents to the online survey were asked to rank the importance of search options used to look for publications, author, title, keywords, abstract and subject searches stood out as the most valuable to respondents. However, in order to fulfil all of those requirements, a full text version of the document stored in the repository is needed in addition to quality metadata. If a full text version is unavailable, then the text of an abstract would be the next best substitute. Work Package 3 has already set targets for the minimum number of full text items to be added to EO.

An element which was highlighted as being important by the User Requirement Report was the creation of a citation index. At the time EO is launched, it will provide rankings of results according to the number of full text downloads. However, during the focus groups which were conducted it became evident that there is an on-going debate on the meaning of these indicators. Many economics researchers think that citations are a measure of 'academic impact' whereas the number of downloads or views of abstracts might measure instead the overall 'popularity' of a paper. To most focus group participants, the first form of measurement was more important. It is recommended that a citation index be considered as part of the Business and Sustainability Plan to further develop the service after the end of the Project.

#### **4.4. Multilingual functionality**

The current work plan states that the central gateway will include multilingual functionality so as to respond to the linguistic diversity that characterises European



research. Work Package 6 will develop a search interface in English, French, German and Spanish, with the possibility to increase the number of languages supported in the future. Additionally, the metadata aggregated from the local repositories is planned to be searchable in the four languages stated above by means of a translation tool.

The information presented in the User Requirement Report indicates that economics researchers predominantly speak English (98%) and use this for producing academic publications. The multilingual functionality may therefore, constitute a complementary tool for those who conduct specialised searches. Recommendations from the online survey for the priority of the additional three languages to English varied according to region (for example, respondents from France, Belgium and the UK & Ireland favoured French after English; those from Germany, the Czech Republic and the Netherlands favoured German; economists from Spain favoured Spanish). It is recommended that this functionality should be able to be customised according to the end user.

#### **4.5. Additional Services**

The current work plan includes the possibility of providing a series of added-value services such as: email alerts; the collection and communication of usage statistics like downloads of documents to authors; access to the publisher's version with the implementation of Open URL; links to institutional Inter-Library-Loan (ILL) services; and searches for the complete biography of co-author(s).

Most of the added value services considered by the EO work plan were considered useful and necessary by those consulted for the User Requirement Report. The provision of usage statistics to contributors, for example, would constitute an important service, considering that only a minority of respondents (37%) to the online survey indicated having access to information on the use of their publications.

However, it has not yet been thoroughly assessed as to how best to communicate such information as well as how to provide services such as updates of new publications added to EO. Focus group participants noted that they were receiving too many emails already and worried that email alerts and RSS feeds would further contribute to this overload. Therefore, it is recommended that content updates and





usage statistics are not only communicated through optional registration for RSS feeds and email alerts.

There was no clear consensus among economists as to which additional features they would like to be included in EO. Online survey respondents identified some services already specified in the current work plan such as links to local ILL systems and links to the fee-based versions of publications. However, some focus group participants were also keen to incorporate others services which have not yet been considered for inclusion. These included: providing a space within the EO portal to advertise conferences and job openings with emphasis in European institutions so as to increase the visibility of their respective institutions; having a space for discussion and debate in order to assess the quality of papers uploaded (particularly non-peer-reviewed literature such as working papers); being able to tag and comment items within EO. It is advisable, therefore, that these additional services are looked at in the future Business and Sustainability Plan.

A request which received widespread support from economics researchers and is already included in the work plan is to make it possible to export search results into reference management software packages such as EndNote, Reference Manager or Refworks as well as BibTex.

Finally, and as identified in the work plan, increasing the visibility of economics research by enabling additional access to online search engines such as Google (Scholar) and interoperability with RePEc is regarded as very important by potential end users. When consulted, online survey respondents identified Google (64.7%), Google Scholar (69.8%), RePec (66.8%) and SSRN (56.1%) as being the most preferred services for the dissemination of their own research. It is therefore essential that all institutional repositories' metadata on research outputs are easily mapped into the formats defined by Work Package 5 so they can be disseminated to these services.

**Table 2. Summary of service specifications, responsibility for these and status in the work plan**

The below table summarises the service specifications from the Services section within the Report on content acquisition, dissemination and service specifications. The table identifies whether the responsibility for each service specification is at the local repository level (i.e. institutional) or the gateway level (i.e. Economists Online).



There are four statuses and the status of each recommendation has been indicated in the status column. These are:

- *Currently in the work plan*: this indicates that the recommendation has already been envisaged as part of the current work plan.
- *Yet to consider in the current work plan*: this indicates that the recommendation will be discussed by the NEE0 partners at the Project meeting in month 10 (June 2008).
- *Yet to consider in the Business & Sustainability plan*: this indicates that the recommendation is out of the scope of the NEE0 project but will be considered as part of the Business and Sustainability plan. The first version of this is due in month 15.
- *Potential conflict with the current work plan*: this indicates that the recommendation is in potential conflict with the current work plan.

Recommendations	Responsibility	Status
<b>4.1 Services: Metadata ingest</b>		
a. Institutional repositories configured to provide compliant metadata and usage statistics on relevant textual publications to EO.	<i>Institution</i>	<i>Currently in the work plan</i>
b. EO able to harvest compliant metadata on textual publications including usage data.	<i>EO</i>	<i>Currently in the work plan</i>
c. EO able to full text index the following textual publications: <ul style="list-style-type: none"> <li>• PDF: application/pdf</li> <li>• ODT : application/vnd.oasis.opendocument.text</li> <li>• TXT: text/plain</li> <li>• HTML: text/html</li> <li>• [La]TeX: application/x-latex</li> <li>• PostScript: application/postscript</li> <li>• MS-Word: application/msword</li> </ul>	<i>EO</i>	<i>Currently in the work plan</i>
d. Datasets to be described using DDI	<i>Institution</i>	<i>Currently in the work plan</i>
e. Institutional repositories to support Excel, CSV (comma separated), Strata and SPSS file formats for datasets.	<i>Institution</i>	<i>Currently in the work plan</i>
f. EO able to full text index PowerPoint files.	<i>EO</i>	<i>Yet to consider in the current work plan.</i>
g. EO to aggregate metadata from SSRN, if information is made accessible by SSRN.	<i>EO</i>	<i>Yet to consider in the Business &amp; Sust. plan.</i>



Recommendations	Responsibility	Status
<p>h. Repositories should have a flexible submission process. Scenarios which could be considered:</p> <ul style="list-style-type: none"> <li>i. harvesting of published works in a semi or automatic manner</li> <li>ii. linking repositories directly into institutions' content management systems so as to enable automatic capture at the time of first publication.</li> </ul>	<i>Institution</i>	<i>Yet to consider in the Business &amp; Sust. plan.</i>
<b>4.2 Services: Datasets</b>		
<p>a. Institutional repositories configured to support ingest and storage of datasets and associated metadata including cross links to relevant textual publications.</p>	<i>Institution</i>	<i>Currently in the work plan</i>
<p>b. Metadata scheme to include ability to store specific statements on creator's rights and end-user permissions. NB May also be a requirement to be hold a reference to original dataset owners rights where working dataset is partly derived from commercial data. Complex rights requirements which require further expert consideration.</p>	<i>Institution</i>	<i>Currently in the work plan</i>
<p>c. Institutional repositories or data archives to restrict access to users who have been presented with and have agreed to respect creator's rights and abide by associated statement of end-users permissions.</p>	<i>Institution</i>	<i>Currently in the work plan</i>
<p>d. EO able to harvest compliant metadata for datasets including usage data.</p>	<i>EO</i>	<i>Currently in the work plan</i>
<p>e. Institutional repositories or data archives configured to support ingest and storage of associated tools such as command or syntax files as well as technical software extensions and/or applications developed by researchers with innovative methodologies for data analysis.</p>	<i>Institution</i>	<i>Yet to consider in the current work plan.</i>
<b>4.3 Services: Searching</b>		
<p>a. Full text searching of the following textual publications &amp; associated metadata.</p> <ul style="list-style-type: none"> <li>• PDF: application/pdf</li> <li>• ODT : application/vnd.oasis.opendocument.text</li> <li>• TXT: text/plain</li> <li>• HTML: text/html</li> <li>• [La]TeX: application/x-latex</li> <li>• PostScript: application/postscript</li> <li>• MS-Word: application/msword</li> </ul>	<i>EO</i>	<i>Currently in the work plan</i>
<p>b. Metadata search by: author, title, keyword, abstract, subject.</p>	<i>EO</i>	<i>Currently in the work plan</i>
<p>c. Limit searches by: date, institution, JEL codes, document type</p>	<i>EO</i>	<i>Currently in the work plan</i>

Recommendations	Responsibility	Status
d. Option to rank results by number of full text accesses	EO	<i>Currently in the work plan</i>
e. Option to sort results by date.	EO	<i>Currently in the work plan</i>
f. Full text searching of PowerPoint files and associated metadata.	EO	<i>Yet to consider in the current work plan</i>
g. Option to rank results by the number of times item it has been cited and a facility to display list of citing articles.	EO	<i>Yet to consider in the current work plan.</i>
h. Option to sort results by author and title.	EO	<i>Yet to consider in the current work plan.</i>
i. Option to specify preferred combination of ranking/sorting options.	EO	<i>Yet to consider in the current work plan.</i>
j. EO able to recognise citations in textual documents and consequently maintain citation counts for all textual documents.	EO	<i>Yet to consider in the Business &amp; Sust. plan.</i>
<b>4.4 Services: Multilingual functionality</b>	EO	
a. End users to be able to choose multilingual functionality from English, German, French and Spanish.		<i>Currently in the work plan</i>
b. User reports indicates that there is little requirement to provide language support beyond English.	EO	<i>Potential conflict with the current work plan</i>
<b>4.5 Services: Additional services</b>	EO	
a. Provision of statistics on full text accesses of textual documents to authors.	EO	<i>Currently in the work plan</i>
b. Bibliographic references exportable to EndNote and Reference Management reference management software packages.	EO	<i>Currently in the work plan</i>
c. EO textual publications and datasets visible and discoverable via Google and Google Scholar	Institution	<i>Currently in the work plan</i>
d. EO metadata to be uploaded into RePEc.	EO	<i>Currently in the work plan</i>
e. Metadata configured through Open URL resolver of institution to hold links to final published versions of textual documents including links to external commercial services.	Institution	<i>Currently in the work plan</i>
f. Provision of statistics on downloads of datasets and associated tools to authors.	EO	<i>Yet to consider in the current work plan.</i>
g. Provision of statistics on the number of times textual documents are cited to authors.	EO	<i>Yet to consider in the current work plan.</i>
h. Email alerts and RSS feeds should be optional and customisable.	EO	<i>Yet to consider in the current work plan.</i>
i. Incorporate a forum into EO. This could be done on the website.	EO	<i>Yet to consider in the current work plan.</i>



Recommendations	Responsibility	Status
j. Ability to tag and comment on items in EO.	EO	<i>Yet to consider in the current work plan.</i>
k. Bibliographic references exportable to BibTex reference management software package.	EO	<i>Yet to consider in the current work plan</i>
l. EO metadata uploaded into SSRN.	EO	<i>Yet to consider in the current work plan</i>
m. Conference announcements and job openings to be advertised on the EO website	EO	<i>Yet to consider in the Business &amp; Sust. plan.</i>



## References

Pauwels B. (2008) *NEEO Technical Guidelines* (Draft – version 2.2). NEE0.

Place T. (2008) *1<sup>st</sup> report on the SOA architecture design* (Report). NEE0.

Reid M. (2008) *First report on IPR Issues* (Report) London: NEE0.

Torres CA, Blake M, Shipsey FM. (2008) *User requirement report* (Report). London: NEE0.

Directorate - General Information Society and Media (2007) *Grant agreement Number ECP-2006-DILI-510001 Annex 1. Description of Work (Targeted projects)*. Luxembourg: Commission of the European Communities.