

e-Research: e-Infrastructure Proposal Cover Sheet

Cover Sheet for Proposals (All sections must be completed)	JISC Capital Programme
--	-------------------------------

Name of Capital Programme: e-Research: e-Infrastructure	
Bid for Call: (Please tick ONE BOX ONLY, as appropriate)	
Text Mining	
<input checked="" type="checkbox"/>	Call I – Automatic Summarisation for Systematic Reviews using Text Mining
Name of Lead Institution: Institute of Education, University of London	
Name of Proposed Project: Text mining as a search engine for the UK Educational Evidence Portal	
Name(s) of Project Partner(s): The UK Educational Evidence Portal Consortium, supported by CfBT Education Trust	
Full Contact Details for Primary Contact:	
Name: James Thomas	
Position: Senior Research Officer	
Email: j.thomas@ioe.ac.uk	
Address: EPPI-Centre, Social Science Research Unit, Institute of Education, 18 Woburn Square, London, WC1H 0NR	
Tel: 020 7612 6844	
Fax: 020 7612 6400	
Length of Project: 18 Months	
Project Start Date: 1 st October 2007	Project End Date: 31 st March 2009
Total Funding Requested from JISC: £113,775 (+ £60,000 to NaCTeM = £173,775)	
Funding Broken Down over Financial Years (Mar–Apr):	
Apr 07 – Mar 08	Apr 08 – Mar 09
£37,053.30 (+ proportion of NaCTeM)	£76,721.90 (+ proportion of NaCTeM)
Total Institutional Contributions: nil	
Text mining, that interrogates and sorts thousands of electronic documents, has the potential to revolutionise the identification of research and collective expertise for policy-making and practice. This proposed project will develop and test text mining to support efficient access to evidence through a web portal. Currently, access to evidence is either available through general web search engines or specialist bibliographic databases. The scope of bibliographic databases is often narrow, necessitating multiple searches, whereas the scope of a search engine such as Google is very broad, resulting in huge numbers of irrelevant hits. We therefore propose to evaluate the extent to which text mining is able to overcome these problems in a case study of the UK Educational Evidence Portal. This proposal is for the Evidence for Policy and Practice Information and Co-ordinating Centre (specifically its user involvement team) to work with the National Centre for Text Mining to develop and evaluate an innovative search engine – using text mining – for a portal of education evidence, relevant to education practitioners and policy-makers. If successful, this project would be a high profile exemplar of the utility of text mining in the social sciences, with application beyond the single case described here.	
I have looked at the example FOI form at Appendix A and included an FOI form in the attached bid (Tick Box)	<input checked="" type="checkbox"/>
I have read the Circular and associated Terms and Conditions of Grant at Appendix B (Tick Box)	<input checked="" type="checkbox"/>

FOI Withheld Information Form

We would like JISC to consider withholding the following sections or paragraphs from disclosure, should the contents of this proposal be requested under the Freedom of Information Act, or if we are successful in our bid for funding and our project proposal is made available on JISC's website.

We acknowledge that the FOI Withheld Information Form is of indicative value only and that JISC may nevertheless be obliged to disclose this information in accordance with the requirements of the Act. We acknowledge that the final decision on disclosure rests with JISC.

Section / Paragraph No.	Relevant exemption from disclosure under FOI	Justification
NONE		

Please see <http://www.ico.gov.uk> for further information on the Freedom of Information Act and the exemptions to disclosure it contains.

1. Appropriateness and Fit to Programme Objectives and Overall Value to the JISC Community

1.1 Introduction

Text mining, that interrogates and sorts thousands of electronic documents, has the potential to revolutionise the identification of research and collective expertise for policy-making and practice. This proposed project will develop and test text mining to support efficient access to evidence through a web portal. Currently, access to evidence is either available through general search engines (such as Google) or specialist bibliographic databases. The scope of bibliographic databases is often narrow (and commercially restricted), necessitating multiple searches, whereas the scope of a search engine such as Google is very broad, resulting in huge numbers of irrelevant hits. While initiatives such as the Semantic Web¹ offer the prospect of more efficient searching in the future, efficient present day searching remains problematic. We therefore propose to evaluate the extent to which text mining is able to overcome these problems in a case study of the UK Educational Evidence Portal (EEP). This proposal is for the Evidence for Policy and Practice Information and Co-ordinating Centre to work with the National Centre for Text Mining (NaCTeM) to develop and evaluate an innovative search engine – using text mining – for a portal of education evidence, relevant to education practitioners and policy-makers. If successful, this project would be a high profile exemplar of the utility of text mining in the social sciences, with application beyond the single case described here.

1.2. Aims and objectives

This project aims to demonstrate the applicability of text mining in an emerging education web portal. Specifically, it will:

1. ascertain user needs and priorities for the web portal;
2. develop modified ASSERT tools to operate in the portal;
3. evaluate the portal within its user community;
4. modify the portal in response to user feedback;
5. be an on-going demonstration of how text mining can facilitate the location of evidence to inform users within a social science discipline.

1.3 The UK Educational Evidence Portal

In education, as in many other professions, the internet is being harnessed to provide access to the evidence base underpinning practice and policy making. Recognising that research findings are dispersed across multiple sources and that both professionals and lay people have limited time to seek them out, a consortium of organisations has come together to develop an electronic portal. The Educational Evidence Portal (EEP) draws on a wide range of documents from selected sources and makes them accessible to users via the Web.

EEP is a “bottom-up” initiative driven by a voluntary consortium of organisations and programmes chaired by Andrew Morris, supported by the CfBT Education Trust. It has grown out of an idea developed as part of the programme of the DfES-funded National Educational Research Forum. In order to take the project forward, a group of ‘partner’ organisations has been formed, including: the Department for Education and Skills (DfES), the Training and Development Agency for Schools (TDA), the National Foundation for Educational Research (NFER), the Higher Education Academy (HEA), CfBT Education Trust (CfBT), the ESRC Teaching and Learning Research Programme (TLRP), the British Educational Communications and Technology Agency (Becta) and the EPPI-Centre. Membership of the consortium is open, and has been growing since its formation in 2005. The vision for EEP is that it will make available all online educational evidence that can assist practitioners and policy-makers in their decision-making.

Since 2006, the portal has been developed in conjunction with the Microsoft Corporation as part of its ‘Partners In Learning’ programme with the TDA. The outcome of this work, the second pilot portal, is available online at: <http://www.eep.ac.uk>. The portal is currently hosted by the EPPI-Centre. It is a small web site with a search engine that enables users to search the contents of the websites of 23 organisations (the search for ‘all resources’). As well as the websites of the above partner organisations, the search includes resources such as the General Teaching Council for England, The Higher Education Academy, HM Inspectorate of Education, Ofsted and the Scottish Executive.

¹ <http://www.w3.org/2001/sw/>

A 'selected resources' search enables users to search a much smaller pool of materials that have been assigned terms according to the British Education Thesaurus. At the time of writing this trial search contains approximately 400 items – many orders of magnitude smaller than the contents of the search for all resources – and therefore lacking important evidence. Further developments are under discussion with the British Education Index, which is tagging these selected resources with the British Education Thesaurus and making them available to the EEP portal, so this number will increase over time.

1.4 The current search engine

The current search engine for all resources utilises Microsoft's 'Windows Live' search – similar to Google – via a web service. This is a free-text search that is filtered according to the contents of the 23 specified websites. This search is useful, in that one is able to search the websites of key educational institutions in the UK, but restricted, because it has limited configurability. Notably, it does not enable users to search according to a taxonomy of controlled terms, and also searches areas of members' websites that should not be included in the search (for example, a search for music in schools might also retrieve job vacancies as well as published research). Thus, while the pool of data being searched is reasonable, since it is limited to the websites selected by the consortium, it lacks specificity, mainly due to its dependence on the search terms entered by the user.

1.5 Using meta-tagging to improve the current search

One option to improve the specificity of searches currently being explored by the consortium is the consistent meta-tagging (or 'keywording') of all documents in the portal according to a standard thesaurus. This is, however, not a trivial task. There are currently 23 organisations in the consortium with thousands of documents to be included in the EEP search. Each document would need to be read and have thesaurus terms added to it. Each website would also need to be updated and the terms included as HTML 'META' tags on each page. The task thus requires the time and expertise necessary to assign thesaurus terms, and resources to modify current websites. Moreover, there is no agreed standard with which to categorise educational research in the UK. While there are possible candidates, such as the BEI-maintained British Education Thesaurus (BET), the EEP consortium has not identified a tool that is ideal for use by education practitioners and other potential users. It is however, working with the British Education Index to tag records according to the BET for the 'selected resources' search. (The availability of a set of keyworded documents will provide the proposed project with a useful comparison of two methods of searching: evaluating the difference between utilising manually-applied keywords and text mining.)

Thus, given the current limitations in institutional resources and candidate meta-tagging tools, and the difficulties inherent in persuading multiple organisations to commit significant resources to an enterprise such as this, it is difficult to see how all the documents might be tagged consistently and their meta-tags exposed in a standardised way. In time, meta-tagging frameworks may become commonplace, and the EEP consortium is committed to continue to explore this option; however, it is also sensible to explore alternative mechanisms for improving the specificity of searches.

1.6 Utilising the ASSERT text mining tools as a search engine for EEP

Text mining offers the possibility of improving the specificity of searches without the need to go back and apply meta-tags manually to many thousands of documents. The application of meta-tags to documents creates, *deductively*, 'clusters' of documents around each term. Text mining can also work with clusters of documents, *inductively* deriving conceptual clusters based on the text each document contains. The ASSERT project contains a clustering engine which is currently being tested within systematic reviews of social science literature. The aim of this project would be to test the applicability of using this clustering engine to form the basis of searches in EEP. If it is able to return useful results to the user, the need for retrospective tagging of documents will diminish. In addition, the work in ASSERT relating to query expansion will also be incorporated into EEP.

At the 'back end' of the proposed project, a webcrawler would be used to retrieve all the relevant web pages, pdf files and .doc files from consortium organisations and store them in a database (NaCTeM has tools which can convert these common file types to a suitable format for use in text mining). This database would be indexed by the text mining engine.

There are three versions of the 'front end' user interface that would be evaluated in this project:

1. the user enters their own free-text term(s) and views the results;
2. the user enters their own free-text term and is then able to explore the clusters generated around this term;

3. the user selects one or more terms from a hierarchical taxonomy; these terms having already been used to classify documents into predefined clusters.

Using option 1, users would simply enter the search term(s) of their choice, using a familiar Google-style interface. In the background, the text mining engine would cluster the documents according to this term and return the most relevant results. It is possible that the text mining engine would also identify synonyms and term variants to enhance the search using other external resources such as thesauri or ontologies ('query expansion'). Users would then be able to browse the results and click through to view the original documents.

Using option 2, users would enter their own free-text term and would then be presented with a list of clusters in that area (using the 'Lingo' engine within ASSERT). They would then be able to select the most relevant cluster(s) and explore the documents they contain.

Option 3 makes use of a hierarchical taxonomy of educational research that is being developed by the EEP consortium. The taxonomy will contain a maximum of ten top level areas which each expand to up to ten sub-areas which then expand again to the most detailed level. The taxonomy will therefore contain up to 1000 terms. The clustering engine would operate in a similar way to Option 1 in this model, but the user would have a structured list of terms to aid them in their search. Since the search terms are known in advance, it would be possible to evaluate the utility of using term variants and synonyms to enhance the search beyond simply using the terms in the taxonomy. Moreover, given that the search terms are known in advance, the documents can be clustered in full, without the time constraints inherent in training classifiers for clusters in real time. This should produce better results.

While requiring additional development, the proposed project therefore builds directly on the work already accomplished in the ASSERT project. Document clustering forms the most significant part of the proposed search engine – and this is also at the heart of ASSERT. This project would also build on the work currently underway in ASSERT regarding its ability to expand queries around a given term. However, it is recognised that the three options outlined above are ambitious for a small project, and the search available using Option 3 (above) may not involve significantly more than the engine described in Option 1 being employed using a set of pre-defined terms. In spite of this, the evaluation of such a search would be a valuable exercise, and the project thus has scope to develop and refine the sophistication of the text mining engine.

The only part of ASSERT on which this project does not propose to build is document summarisation. While the summarisation of documents within particular clusters would be a valuable addition to the portal, it is unlikely that this degree of development would be possible in this case study. If the text mining engine can be developed and evaluated favourably, the next developmental step would be to incorporate document summarisation.

While general search engines, such as Google and Windows Live are useful ways into the mass of information available online, users often find that they need to wade through many irrelevant pages before finding the relevant ones. Resources, such as *Roll your own search engine*² have been created in response to this need, but searches using these engines are still not as specific as those possible using meta-tagged materials. The proposed case study is therefore a specific example of a general need. The software tools that arise from this project are designed to facilitate the production of other portals in other topic areas. All that need change is the contents of the back-end database. While the educational taxonomy described in Option 3 would not necessarily be relevant to other contexts, the utility of using a taxonomy, and possibly tailoring the term variants and synonyms used, will be applicable to fields beyond education alone. The project is therefore a good match with the call for funding: it is a case study arising from a genuine user need with substantial user involvement that is capable of raising the profile of text mining in the social science community. In view of this wider applicability, and cognisant of the requirements in the call, time has been set aside in the project plan to engage with the wider JISC community in publicising the potential of utilising text mining as described in this project.

The proposed project is innovative on two levels. First, it proposes to utilise text mining on a live search engine that has the potential to serve thousands of practitioner users. Second, it is an innovative solution to a problem – that of hand-tagging thousands of pre-existing documents – that has application beyond the proposed case study. Thus, if successful, this project would not only provide a high profile example of text mining working in a social science discipline with significant benefits beyond the JISC and academic community, it would also provide a possible way forward for other projects, which have similar needs in terms of making large numbers of untagged documents available for effective searching.

² <http://www.rollyo.com/>

1.7 Project outline

There are two distinct areas of work inherent in this project, split between NaCTeM and the EPPI-Centre. As outlined in the call, it is anticipated that NaCTeM would be responsible for the development of the modified ASSERT tools needed to drive the search engine. The EPPI-Centre would be responsible for the requirements analysis, liaison with the EEP consortium and evaluation of the final product.

There are seven distinct stages of work in this project:

1. requirements analysis;
2. development of the web crawler and database;
3. modification of the ASSERT tools to produce the search engine;
4. initial evaluation of the search engine;
5. initial modification of the search engine;
6. final evaluation of the search engine;
7. final modifications in the light of the above evaluation.

Since the modifications to ASSERT would be the responsibility of NaCTeM, this proposal concentrates on the methods to be used for the requirements analysis and evaluation.

1.8 Evaluation of text mining for EEP

The EEP portal is designed for both professional and lay people interested in education, to help them find educational evidence from a range of reputable sources using a single search. Nevertheless, little is known about what sort of evidence users are actually seeking, whether they find it and, if so, recognise its relevance and utility to their own work. Text mining through the EEP portal has the potential to help them navigate thousands of documents and collate those most relevant to their query. Whether this potential is achieved will be determined by addressing the following evaluation questions.

Compared with the available current EEP search strategies of free text or keyword / thesaurus searching:

1. Does text mining increase the output of documents relevant to users' queries?
2. Does text mining deliver the output of relevant documents more quickly? (i.e. do users locate relevant information more quickly?)
3. Does text mining guide the thinking of users interrogating EEP for documents? (Particularly with regard to document clustering.)
4. What is needed to improve users' access, comprehension and use of documents?

Utilising the EPPI-Centre's combined expertise in systematic coding of evidence, and in user involvement in research, the evaluation will be in three stages: (1) a formative evaluation of users' needs and responses to the portal, through focus groups and an on-line questionnaire which will inform the requirements analysis; (2) a comparison of the outputs of text mining, free text and keyword / thesaurus searching in terms of sensitivity, specificity and clustering for documents providing different types of evidence across a range of topics; (3) a pilot study of users searching EEP for their own interests.

Assessing users' needs for the requirements analysis

Drawing on our experience of working alongside service users, initially we shall assess users' needs for evidence, and their need for guidance for accessing that evidence through the portal.

Potential users will be sought by asking EEP consortium members to identify types of people who use their websites and other services, and suggesting forums for convening focus groups, such as teachers in schools, policy makers in government departments, or youth leaders in charities. The EEP consortium is well-placed to identify and recruit a diverse range of potential users. A maximum of six focus groups of 5-7 people each, will explore their evidence needs, introduce them to EEP on-line and seek their expectations and initial responses to the content and interface. Participants will be sent an internet link to the portal a week before the focus group meeting, and will be asked to note their impressions, likes and dislikes. During the focus groups we shall ask participants to discuss their impressions of the portal with particular focus on the interface and the relevance of the information and the degree to which it is understandable. Particular attention will be paid to the topics and questions for which potential users might search the portal, and their awareness and understanding of the different types of evidence that might be sought, and their own success criteria for portal searching. Two members of staff will attend the focus groups: one to observe and the other to facilitate. The findings of these focus groups will inform the requirements analysis that will be submitted to NaCTeM to form the basis of their modifications of the ASSERT tools.

Comparative study of text mining

Following the modification of the ASSERT tools and the development of a useable text mining search engine, the research team will search the portal for topics and questions raised by members of focus groups using (a) the current 'Windows Live' free text search, (b) keywords from the British Education Thesaurus, and (c) the searches available using text mining. They will compare the sensitivity, specificity and clustering of these methods and how they perform in relation to users' success criteria. The findings will inform the design of help pages to guide users in navigating the portal.

The results of this phase of the evaluation will feed into the initial modifications of the search engine (Stage 5 in the Project Outline).

User testing

We shall return to EEP members to recruit a maximum of 30 individuals from a range of user groups for testing the developments. User testing will address the portal interface, the guidance developed to support searching by users, and users' recognition of the relevance and utility of the documents retrieved.

Following the software development, and the development of help pages, we shall invite users to spend an hour to an hour and a half testing the portal. Participants will be given vouchers in acknowledgement of this significant time commitment. Participants will sit one at a time in front of a computer with a test coordinator and be invited to browse the portal pages. The test coordinator will ensure that all elements of the portal are visited. Participants will be asked to describe what they could see on the screen, what they are looking for, and what they like or dislike. They will be asked to identify a topic and a particular question relevant to their interests, and search the portal using the different search options outlined above. With the participant's permission, their use of the portal will be observed and notes made about the participant's expectations of the site, the degree to which they found the information relevant and were able to understand the information and the navigability of the portal and associated sites.

The results from this phase of the evaluation will determine any changes necessary to the search engine, help pages and other aspects of the interface, before the final version of the software is produced (Stage 7 in the Project Outline).

1.9 Exit/Sustainability Plan

The EEP consortium has been working to develop an education portal for two years and now has a pilot site online.³ If, after evaluation and modification, the text mining search proves to be a useful addition to the site, the engine will be incorporated into the current search options – either replacing the search mechanisms currently available, or adding another option for searching from which users can choose. Since the portal has the commitment, and active support, of the major educational organisations in the UK, the portal will be widely publicised – and with it, the innovative use of text mining in its search.

In addition to the potential benefits to users of the portal, the project will be written up as a case study in using text mining tools within a specific social science discipline. The case study would be published in a relevant peer reviewed journal and the project publicised via conferences and appropriate JISC events, and on EEP and consortium websites. This should ensure that learning from the project reaches the appropriate audiences. In line with current NaCTeM practice, the text mining tools themselves will be made available freely for others to use.

1.10 Ethics

Ethics approval will be sought from the Faculty ethics committee at the Institute of Education. Participants will be informed of the purpose and process of the study, their right to withdraw at any time, and the security procedures for storing data.

2. Quality of Proposal and Robustness of Workplan

The proposed programme of work is detailed in the table below. The partner responsible for each work package is identified in the first column.

³ See www.eep.ac.uk

Work package (WP)	Month	Description	Deliverables	Effort (person months)
WP1 EPPI	1-4	Initial meetings between the NaCTeM and EPPI-Centre teams; liaison between EPPI-Centre and EEP consortium; assessment of user needs and development of requirements analysis.	D1 (M1) Specification of requirements of text mining engine and portal front-end.	EPPI 2
WP2 EPPI	2-3	Identification of the sites to be indexed In consultation with the EEP consortium.	D2.1 (M2) List of URLs to be indexed	EPPI 0.25
WP3 EPPI	2-3	Construction of the webcrawler and database.	D2.2 (M3) Database of text items for indexing	EPPI 0.5
WP4 NaCTeM	4-9	Main modifications to the ASSERT tools.	D3.2 (M4) Text mining engine and web services online	NaCTeM 5
WP5 EPPI-	8-9	Construction of the html front end of the portal and incorporating interfaces to the 3 search options.	D3.2 (M5) The front end of the portal	EPPI 0.5
WP6 EPPI	10	The 'comparative study of text mining': EPPI-Centre evaluation based on the requirements gathered from users.	D4 (M6) Document specifying the results of this evaluation and recommendations for modification before user testing.	EPPI 1
WP7 NaCTeM	11	Minor modifications to the text mining engine	D5 (M7) Modified text mining engine	NaCTeM 1
WP8 EPPI	12-15	Detailed user testing of the three search options.	D6 (M8) Detailed evaluation of user experiences using the 3 search options; recommendations for modification.	EPPI 2
NaCTeM	16	Modifications to the text mining engine in the light of user evaluation	D7 (M9) Modified text mining engine	NaCTeM 1
WP9 EPPI	16-18	Final documentation of project processes etc; documentation for downloadable software; academic paper.	D8 (M10) Completed documented processes and workflows; academic paper.	EPPI 2 NaCTeM
WP10 EPPI	1-18	Engagement with JISC programme activities		EPPI up to 0.5
WP11 EPPI	1-18	Project management	Facilitates the above deliverables	EPPI PD and co Is

The estimates of effort and work to be undertaken by NaCTeM are indicative; they may change after the requirements phase is complete and we have a detailed specification for the modification work. (Since the current portal is hosted by the EPPI-Centre, it is proposed that the construction of the database and interface to the text mining tools is carried out by the EPPI-Centre. However, after requirements are examined in detail, it might be more practicable for this development to take place at NaCTeM.)

2.1 Deliverables

The main deliverables are:

- The specification of requirements (D1), which consists of the results of the initial focus groups.
- The initial (D3.2) and final (D7) versions of the modified ASSERT tools. This will demonstrate the applicability of text mining tools in this area of social science. In the event of the tools not meeting the needs of the EEP user community, and therefore not being incorporated into the live portal, the tools will still remain online as a demonstration of a possible way forward in the future. As mentioned above, the modified ASSERT tools will also be available for other users.
- The detailed user evaluation of the portal (D6) which will inform development of the final version.
- The documented processes and workflows (D8) and detailed description of the case study. This deliverable would also contain the user evaluations and recommendations for future development.

2.2 Risks, IPR and sustainability

The main risk to an effective user requirements analysis and the subsequent evaluation is a failure to recruit appropriate users. However, the EEP consortium is well placed to identify and recruit a wide range of potential users – and has a track record of achieving this in the past when developing the current portal. This risk is effectively controlled by the large number of supporting organisations and their access to key user groups.

The risk that the requirements analysis would not be conducted properly – and that, therefore, the portal would be built to the wrong specification must also be considered. This risk is effectively controlled by the fact that the team leading the evaluation has a distinguished track record of developing networks amongst users and undertaking research among user communities.

A common problem with projects that require working across organisations is the development of effective working relationships and good communication. The risk that this problem becomes an issue in this project is controlled, to a large degree, by the fact that the project director has already worked with NaCTeM on ASSERT and is on the EEP consortium Strategy Group. The work involved in establishing new relationships has thus already been accomplished.

IPR: as a matter of course, NaCTeM makes the software it develops available to the widest possible user community. There is no need to deviate from this policy in this case and, indeed, the tools can only benefit from other organisations testing and refining them for their own circumstances.

Sustainability is an issue that is not difficult for this project to address. All the indications are thus far that the text mining tools would be a significant benefit to the EEP portal. If this proves to be the case, the tools would be incorporated into the EEP portal and maintained by the consortium. This would therefore provide a high profile example of text mining facilitating searches in a wide user community. In the unlikely event that the tools do not meet the needs of the users, the pilot portal would continue to be hosted at the EPPI-Centre as a demonstration of possible future ways forward.

2.3 Project management

The project director, James Thomas (0.2 FTE), will be responsible for overseeing the project and overall co-ordination. He will be to be the main point of contact for NaCTeM and EEP, and ensure that the project is characterised by good communications and working relationships between the partners. He will also be responsible for ensuring that the requirements analysis and evaluations are delivered on time and effectively communicated to NaCTeM and for ensuring the pilot portal is published on the EPPI-Centre web server.

Co-investigator Sandy Oliver (0.1 FTE), with her extensive experience of user involvement in research, will lead the requirements analysis (WP1), the initial (WP4) and final (WP6) evaluations of the search engine. She will oversee: the development of the topic guides for focus groups, the analysis of the responses to the portal and expressed interests (topics and questions) and the use of these to test the search capacity of text mining, and the testing of the portal searching by individual users.

Co-investigator Ruth Stewart (0.4 FTE), will draw on her skills and experience of engaging a wide range of users to conduct the requirements analysis, and evaluations with Sandy Oliver. They will share responsibility for facilitating and observing focus groups and individual user testing, and reporting these. Ruth Stewart will draft research tools for working with users, and she will compare the searching capacity of text mining with the use of keywords and free text.

The EPPI-Centre project team would meet regularly to track progress and develop and refine the project plan as required. Members of this team would also meet regularly (in person, or via Access Grid) with the ASSERT team to ensure good communication. The EEP Strategy Group, which meets monthly, would be updated on progress regularly and its advice and input sought. It would thus function as an advisory group for the project, able to give advice from a user perspective.

While FTEs for EPPI-Centre staff have been specified for the duration of the project, it is anticipated that the actual FTEs worked would fluctuate depending on the phase of development. Higher FTEs than specified would be necessary during the requirements analysis and evaluations, while lower FTEs would be needed when NaCTeM was working on the software. This fluctuation is predictable and would form part of the workload planning for individual members of the team.

3. Engagement with the Community

As outlined above, the heart of this proposal, and the strength of the EPPI-Centre team, lies in the engagement of EEP users in the development of requirements and in the final evaluation of the search engine. This is a good opportunity to test the potential for text mining tools to be deployed in a 'real life' situation within the social sciences. As such, the success of the eventual tool rests on its utility to its users and the importance of their views is reflected in the emphasis given to ascertaining their needs and observing their use of the final portal. The EEP consortium too, constitutes an important user group, and the views of this consortium will also feed into the requirements and recommendations for modification. Moreover, the active engagement of this user group is necessary to secure the long-term use of the search engine. As stated under 'exit/sustainability plan', in the event of the evaluation being successful, it is anticipated that the modified ASSERT tools would become a key way that education users would locate evidence to inform their decision-making.

4. Budget

NaCTeM has indicated that it will require a half-time position for between 12 and 15 months to complete the necessary modifications to the ASSERT software. The budget below is for the EPPI-Centre contribution: an additional £60,000 over the lifetime of the project is requested for the NaCTeM portion of work. The 'Other' in non-staff costs covers focus group expenses (vouchers, rooms and travel) and their transcription (£1,500). The relatively small amount of money allocated to 'dissemination' reflects the fact that much of the project is itself dissemination and the fact that staff time has been allocated specifically for this purpose in the workplan.

Non-Staff	April 07– March 08	April 08– March 09	TOTAL £
Travel and expenses	£500.00	£1,000.00	£1,500.00
Hardware/software	£62.00	£0.00	£62.00
Dissemination	£0.00	£200.00	£200.00
Evaluation	£0.00	£0.00	£0.00
Other	£1,350.00	£2,550.00	£3,900.00
Total Directly Incurred Non-Staff (B)	£1,912.00	£3,750.00	£5,662.00
Directly Incurred Total (A+B=C) (C)	1912	3750	5662
Directly Allocated	April 07– March 08	April 08– March 09	TOTAL £
██████████	██████████	██████████	██████████
Estates (RFTE 0.7 Yrs 1 & 2)	£1,700.65	£3,472.23	£5,172.88
Other	£1,848.00	£3,696.00	£5,544.00
Directly Allocated Total (D)	£21,928.10	£45,994.42	£67,922.51
Indirect Costs (RFTE 0.7 Yrs 1 & 2) (E)	£13,213.20	£26,977.48	£40,190.68

Total Project Cost (C+D+E)	£37,053.30	£76,721.90	£113,775.19
Amount Requested from JISC	£37,053.30	£76,721.90	£113,775.19
Institutional Contributions[1]	£	£	£
Percentage Contributions over the life of the project	£113,775.19	Partners	Total: £113,775.19
	X 100%	X %	100%

5. Previous Experience of the Project Team

The Evidence for Policy and Practice Information and Co-ordinating Centre (EPPI-Centre⁴) is part of the Social Science Research Unit at the Institute of Education, University of London. Founded in 1995, the Centre is dedicated to making reliable research findings accessible to the people who need them, whether they are making policy, practice or personal decisions. The project team includes the Centre's research staff with specialism in user perspectives and participation. The Centre is part of the ESRC National Centre for Research Methods, hosting the Node in Methods for Research Synthesis. Taken together, the team outlined below blends experience of the ASSERT and EEP projects; evaluation of user views; software development; and project management.

Dr James Thomas is an Associate Director of the EPPI-Centre and is currently Acting Assistant Director of the Social Science Research Unit at the Institute of Education, London. Since 1994, he has been involved in a wide range of SSRU projects being responsible for: managing, writing and supporting systematic reviews across education and health; methodological development for tools to collect, appraise and synthesise research findings for systematic reviews; and capacity building for the systematic evaluation of educational research. He is also lead developer for EPPI-Reviewer, the EPPI-Centre's software for data extraction and research synthesis and leads a module on methods for research synthesis on the SSRU MSc 'Evidence for Public Policy and Practice'. He is currently Associate Director of the Methods for Research Synthesis node in the ESRC National Centre for Research Methods, leading a stream of work in this programme concerned with e-social science. As part of this work he has been working with NaCTeM on user requirements for the ASSERT project and has been involved in the EEP consortium for two years.

Current Post: Acting Assistant Director, SSRU and Senior Research Officer

Qualifications

PhD (history and social policy): University of London (2005)
MMus (Music): University of Otago (1994)
MA (Music): University of York (1991)
GGSM: Guildhall School of Music (1990)

Major grants

1) National Centre for Research Methods: Methods for Research Synthesis Node (with D. Gough, S. Oliver, A. Harden), £748,113, April 2005 to March 2008, ESRC; 2) RCTs for policy interventions?: a review of reviews and meta-regression (with A.Sowden, I.White, J.Shepherd, J.Dinnes, R. Rees, A.Bagnell, J. Colquitt), £197,068, April 2004 to March 2006, NHS Methodology Programme; 3) Accidental injury, risk taking behaviour and the social circumstances in which young people live: a systematic review (with D. Gough, A Oakley, J. Kavanagh, A. Harden) £179,840, Sep 2004- Dec 2005, Department of Health; 4) Rapid evidence assessments to support HMT policy reviews (with M. Newman), £80,000, May 2006, HM Treasury.

Dr Sandy Oliver is a Reader in Public Policy, Deputy Director of the EPPI-Centre and Deputy Director of the Social Science Research Unit at the Institute of Education, London. Since 1996 she has led a programme of work investigating and meeting people's needs for research based information, working with public service users, practitioners and policy makers – the Perspectives and Participation in Research programme. She

⁴ <http://eppi.ioe.ac.uk/>

has been responsible for designing, managing and reporting primary research projects and systematic reviews; methodological development for data collection and analytical tools investigating responses to research evidence; facilitating discussions of complex research issues with non-researchers; and consensus methods for eliciting collective priorities for research based information. She is currently Deputy Director of the Methods for Research Synthesis node in the ESRC National Centre for Research Methods, leading a stream of work in this programme concerned with research users' involvement in the generation and use of research evidence.

Current Post: Deputy Director, SSRU and Reader in Public Policy

Qualifications

Ph.D (Virology) University of Surrey (1981)
BA Hons. 2(i) (Biology) University of York (1976)

Major Grants

1) Supporting Centres for Systematic Review of Health Policy and Systems Research in Low and Middle Income Countries: World Health Organisation, Alliance for Health Policy and Systems Research, £26,962. April 2007 – March 2008. 2) Randomised controlled trials for policy interventions?: a review of reviews and meta-regression, project funding from the NHS Methodology Programme to Sandy Oliver, Amanda Sowden, James Thomas and others. A multi-centre project collaborating with the Universities of Cambridge, Southampton and York. April 2004 – March 2006. £197,068. 3) Integration and evaluation of consumer involvement in the HTA programme, project funding from National Coordinating Centre for Health Technology Assessment. April 2003 – May 2006. £67,815. 4) ESRC National Centre For Research Methods: Methods for Research Synthesis node. Funded by ESRC to David Gough, Sandy Oliver, James Thomas, Angela Harden. April 2005 – March 2008. £699,272. 5) Realising the research potential of the national newborn bloodspot bank to advance science and health. Funded by MRC, Grant held by Carol Dezateux, Institute of Child Health; Sandy Oliver, Institute of Education; Peter Scrambler, Institute of Child Health, 1 March 2005 – 30 September 2007, £238,352. I lead work stream 1 (of 3), a public consultation with £69,310. 6) Systematic reviews facility in health promotion and public health, funded by Department of Health. Grant held by Ann Oakley, Sandy Oliver and Angela Harden 1 August 2004 – July 2007, Approx £975,000. 7) Newborn Screening Programme Centre, commissioned by the Department of Health. Grant held by Jane Collins, Great Ormond Street; Carol Dezateux, Institute of Child Health; and Sandy Oliver, Institute of Education. I lead the information and training programme for parents and professionals: April 2002 – March 2007, £332,274.

Ruth Stewart has undertaken research with service users, practitioners and policy-makers in health and education, facilitating and researching user-involvement in evidence-informed decision-making for over eight years. As a key member of the EPPI-Centre's Perspectives and Participation in Research programme, her experience includes the use of a wide range of research techniques including conducting interviews and focus groups, developing on-line questionnaires, and employing both qualitative and quantitative analyses. Her background includes undertaking public consultations, conducting needs assessments, and working along side users to develop evidence-informed information and training resources. She has analysed online patient information and worked with users to identify and make sense of electronic resources. As well as designing websites to provide easy access to evidence-based information for patients and practitioners in the area of newborn screening, she has worked alongside NHS Direct library and information services to ensure joined-up provision of information for patients.

She has undertaken research and training in evidence-informed decision-making at practice and policy levels in HIV/AIDS in southern Africa, where the challenges of information access, particularly of online resources, are considerable. In addition, she is a course tutor on the MSc in Evidence-Informed Public Policy at the Institute of Education, London, and is involved in the design and delivery of the module on Personal, Practice and Policy Decision-making.

Current Post: Assistant Director, SSRU and Research Officer

Qualifications

BA (Hons) and MA (CANTAB) Cambridge University (1999)
PhD (education) University of London (Submitted February 2007)