

A. 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	
√	Call I – Automatic Summarisation for Systematic Reviews using Text Mining
Knowledge Organisation and Semantic Services	
	Call II – Semantic Tools for Assisting the Research Lifecycles A) Semantic tools to automate discovery and usage of research resources B) Semantic tools to automate establishment and building research collaborations
Name of Lead Institution:	ESRC National Centre for e-Social Science, University of Manchester
Name of Proposed Project:	Using Text Mining for Frame Analysis of Media Content
Name(s) of Project Partner(s):	ESRC National Centre for e-Social Science (NCeSS) NaCTeM ESRC Centre for Research on Socio-Cultural Change (CRESC), University of Manchester
Full Contact Details for Primary Contact:	
Name: Peter Halfpenny	
Position: Professor	
Email: Peter.Halfpenny@manchester.ac.uk	
Address: Department of Sociology, University of Manchester, Manchester M13 9PL	
Tel: 0161 275 2493	
Fax: 0161 275 1390	
Length of Project:	18 months
Project Start Date:	01/10/07
Project End Date:	31/03/09
Total Funding Requested from JISC: £179992	
Funding Broken Down over Financial Years (Mar–Apr):	
Apr 07 – Mar 08	Apr 08 – Mar 09
£14639	£121438

Total Institutional Contributions:

£44998

Outline Project Description:

Frame analysis has been widely adopted for investigating how texts are framed in a certain way to shape the perceptions or opinions of the information's recipients. It is a more challenging procedure than content analysis in its more quantitative guises, in which the focus is on counting and correlating word frequencies. Although computer-assisted qualitative data analysis (CAQDAS) packages are available to manage and manipulate textual and/or multimedia data, they are not sufficiently advanced to automate the interpretive work of coding that lies at the heart of frame analysis, nor do they support complex retrievals that need to cope with language variability such as synonymy and polysemy. Whilst the amount of social data is growing at an unprecedented speed, a scalable solution for the analysis of large corpora is urgently required.

This project, in collaboration with the ESRC Centre for Research on Socio-Cultural Change (CRESC), aims to customise the ASSERT tools for analysing news texts, and to provide a use case for widening awareness and adoption of text mining to all disciplines in social sciences and humanities. We will 1) investigate frame analysis practices to define initial user requirements for text mining tools; 2) use an iterative process based on design, rapid prototyping, evaluation and refinement, customise the ASSERT suite of text mining tools for the qualitative research community; 3) establish an evaluation framework which may be used in new applications of text mining tools; 4) investigate potential barriers to adoption, sustainability issues and establish responses such as user training and support; 5) document the application of text mining tools for media research as a JISC e-Framework use case.

I have looked at the example FOI form at Appendix A and included an FOI form in the attached bid (Tick Box)

YES
✓

NO

I have read the Circular and associated Terms and Conditions of Grant at Appendix B (Tick Box)

YES
✓

NO

B. 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

Using Text Mining for Frame Analysis of Media Content

PI: Peter Halfpenny

Co-Investigator: Farida Vis

Consultant: Peter Golding

C Introduction

C.1 Background

1. 'Framing' has become a popular research tool, employed widely across a broad sweep of the social sciences and humanities. Frame analysis has been adopted for investigating how texts (including survey questionnaires, media content such as advertisements, news, and tv or radio transmissions, corporate or governmental documents, or our everyday conversations) are framed in a certain way to shape the perceptions or opinions of the information's recipients. There has been much discussion about how a frame should be defined since Goffman established its methodological foundations in 1974¹. Despite continuing discussion and debate, Entman's definition is widely recognised²:

To frame is to select some aspects of a perceived reality and make them more salient in a communicating text, in such a way as to promote a particular problem definition, causal interpretation, moral evaluation, and/or treatment recommendation (Entman, 1993: 52).

2. In accordance with this definition, a 'frame highlights some bits of information about an item that is the subject of a communication, thereby ... making a piece of information more noticeable, meaningful or memorable to audiences' (Entman, 1993: 53). Being able to detect metaphors, symbols, patterns of linguistic composition, or repetition of words in texts is the key to frame analysis.

3. However, the presence of frames in the text is not immediately visible or evident. Framing is an interpretive practice, heavily dependent on the cultural context within which the text is created and consumed. Frames are subtly embedded in texts and their detection relies on teasing out tacit understandings that are normally unconscious resources deployed by readers to make sense of the text. There is no standard, context-independent criterion by which to validate the existence of a particular frame; indeed, a frame detected by one researcher may not necessarily be recognised by another. As Reese remarks, he found it difficult teaching frame analysis to 'those who do not share all of his linguistic and cultural backgrounds' (Reese, 2007: 151)³. Frame analysis, culturally coloured, is thus a more challenging procedure than content analysis in its more quantitative guises, in which the focus is on counting and correlating word frequencies.

4. Frames, being latent in the text, are hard to capture and summarise. To find supporting evidence, researchers usually read texts piece by piece, sorting out textual elements manually, classifying them into one frame category or another, and tagging them with the selected category's name (or code). This is a laborious and time-consuming task, even with the use of current computer-assisted qualitative data analysis (CAQDAS) packages to assist in tagging the text elements. These packages are misnamed because, although they help in retrieving and organising tagged text, they rely on the human operator for the analysis of the text, that is, the decision about which tag (or code) to apply to each segment. The labour-intensity of frame analysis is compounded by the huge and increasing amount of textual information that researchers now have available in the media (including print, such as newspapers and magazines, transcripts of radio and tv transmissions, and narratives on the internet). The individual analyst faces a daunting task of tagging all this material alone or of recruiting a team of analysts, in which case the possibility of the discovery of different frames mentioned above obtrudes: inter-coder reliability is notoriously low⁴.

1 Goffman, E. (1974). *Frame analysis*. New York: Free Press.

2 Entman, R. M. (1993). Framing: Toward clarification of a fractured paradigm. *Journal of Communication* 43(4), 51-58.

3 Reese, S. D. (2007). The framing project: A bridging model for media research revisited. *Journal of Communication* 57: 148-154.

4 Lombard, M., Snyder-Duch, J., & Bracken, C. C. (2002). Content analysis in mass communication: Assessment and

5. Frame analysis therefore faces a variety of methodological challenges: how to efficiently locate and sift large bodies of textual evidence, how to apply coding measures reliably and how to systematically generate frames Can these challenges be usefully met by developing powerful computational approaches for classifying text elements and extracting frames? If so, will the theoretical challenge also be met, of developing machine support to interpret the latent meanings of texts rather than merely producing quantitative summaries of their overt features? (Reese 2007: 151) The goal of this project is to explore whether text mining techniques developed by NaCTeM in the ASSERT project are capable of meeting both the methodological and theoretical challenges that face frame analysis, an important and extensively used approach across the social sciences and humanities.

6. Given the methodological challenge presented above, some social scientists have started to explore the potential of computing power to resolve the difficulties. For example, Peter Golding and Thomas König at Loughborough University have recently undertaken a comparative evaluation of CAQDAS packages for the analysis of media content, particularly of their suitability for textual analysis based on searching and interpretive coding strategies, including frame analysis⁵. The packages they looked at include ATLAS.ti⁶, NVivo⁷, MAXqda⁸, Kwalitan⁹, Qualrus¹⁰, QDA Miner¹¹, HyperRESEARCH¹², and Inforapid¹³.

7. However, as the Golding and König research reveals, although these tools help researchers to manage and manipulate textual and/or multimedia data, they are not sufficiently advanced to assist with its analysis. As noted above, the tools do not remove the human tasks of classifying information and assigning codes (or tags) to it. The sorts of support they provide is to maintain a (sometimes structured) list of the codes as the analyst develops them, together with the ability to drag these codes from the list and drop them on segments of data, or vice versa. It remains the user who interprets the data and applies the codes. Once coding is complete, the packages use them for Boolean searches and the retrieval of text segments. Nevertheless, these existing tools do not automate the interpretive work of coding that lies at the heart of frame analysis.

8. A second problem with existing tools is that their search techniques are not advanced enough to support complex retrievals. Their keyword indexing functions are based on classic information retrieval (IR) techniques. They do not use any linguistic tools to cope with language variability such as synonymy and polysemy which may produce many false positives. Moreover, the indexed terms (keywords) they use do not characterise documents semantically with any degree of sophistication. In other words, existing CAQDAS packages using classic IR methods are not powerful enough for to support the analyst in extracting interpretive frames from the textual data, especially if those frames are complex and multifaceted.

9. Third, existing CAQDAS tools are implemented as desktop systems. As a consequence, they are unable to provide the scalable performance that would be essential for the efficient analysis of large corpora.

10. Apart from these technical and methodological disadvantages, there is also the issue of barriers to adoption of CAQDAS tools. Social scientists may be reluctant to try new technologies for research purposes for a variety of reasons, some practical and others more principled. Most widely-known CAQDAS packages are proprietary and some have substantial licence fees. The

reporting of intercoder reliability. *Human Communication Research* 28: 587-604.

5 König, T. (2004). Routinizing frame analysis through the use of CAQDAS. Paper presented at the Biannual RC-33 Meeting, Amsterdam, August 17-20, 2004.

6 <http://www.atlasti.com/>

7 <http://www.qsrinternational.com/>

8 <http://www.maxqda.com/>

9 <http://www.kwalitan.net/engels/index.html>

10 <http://www.ideaworks.com/qualrus/index.html>

11 <http://www.provalisresearch.com/QDAMiner/QDAMinerDesc.html>

12 <http://www.researchware.com/hr/index.html>

13 <http://www.inforapid.de/html/english.htm>

packages are mostly not interoperable, so coding begun in one cannot be transferred to another. They can have steep learning curves (though support is available from the CAQDAS Networking Project¹⁴). Hence, in exploring the usefulness of text-mining techniques for the social sciences, it will be important to address the issue of engaging users and provide them with easily usable, open source tools and to offer technical support.

C.2 Use Case Preliminary Investigation

11. Analysing news texts commonly proceeds through a number of stages. Researchers will familiar themselves with the data through carefully reading the texts many times. Then they will develop coding categories based on their interpretations of the data. This stage is critical because the codes/categories will serve as indicators of meaningful frames, enabling the researchers to annotate and organise the large volume of narratives. Usually researchers would label the texts (either on the electronic text documents or the print-out), write notes, list ideas, diagram relationships or draw concept maps to help them develop a set of codes. State-of-art of qualitative data analysis software such as ATLAS.ti, NVivo or WordStat assists these processes through, for example, visualising the code frequencies and interrelationships, and counting keywords, but the coding itself remains an entirely human activity.

12. Drawing on these codes/categories, researchers look for special vocabularies that are selected, highlighted, emphasised and presented in a specific way when telling the stories in the media. Researchers code the data iteratively, eliminating, combining and subdividing coding categories until they are satisfied that they have found consistent larger themes that connect codes. Occasionally, researchers will seek to verify their conclusions through having their data independently coded by colleagues, or by seeking independent evidence from other sources or by using other methods such as interviews, focus groups or experiments.

13. Newspapers form the core data for this use case and the substantive topic is how issues around war and conflict, with a particular emphasis on the Middle East are framed. We will explore the various ways in which these issues, Israel/Palestine, Iraq, Iran and so on, are portrayed in mainstream British and American newspapers, and how such issues are framed within a national context. In pursuing this, we will learn how text-mining techniques can contribute to identifying frames and presenting them. We will extend our understanding of whether and how text-mining techniques can overcome the methodological and theoretical challenges faced by frame analysis as it is applied within theoretically-informed empirical media studies research in the social sciences.

14. Our preliminary investigations with CRESC have identified how text-mining tools might benefit their conduct of frame analysis:

- Text mining tools as a generator of frames through extracting information from newspaper texts;
- Text mining tools as a second coder for verifying the frames identified by the researchers;
- Text mining tools for exploring and discovering novel connections and associations between frames identified in newspaper texts.

D Project Description

D.1 Objectives

15. The burgeoning quantity of media texts becoming available to the social science research community and the limitations of the current CAQDAS packages suggest that frame analysis is ripe for development in ways that address the methodological and theoretical concerns identified above. Text-mining technologies offer the opportunity of processing large amounts of textual data systematically, reducing human errors, and saving time. They have the potential to at least partly automate the generation of frames, to a greater extent than current CAQDAS packages.

16. This project aims to illustrate how text-mining technologies might advance frame analysis in social science research. The project has two objectives: 1) customising the ASSERT tools for application to frame analysis of newspaper text; 2) providing a use case to extend awareness

14 <http://caqdas.soc.surrey.ac.uk/>

and promote adoption of text mining across all social science disciplines. To achieve these objectives we will:

- investigate newspaper frame analysis practices to define initial user requirements for text-mining tools;
- use an iterative process of design, rapid prototyping, evaluation and refinement to customise the ASSERT suite of text-mining tools for the media research community;
- develop a generic evaluative methodology to compare the outcome of text-mining with an analysis of the same newspaper dataset that relies on human application of frame analysis;
- use the outcome of the evaluation as a demonstrator of the efficacy of the suite of text-mining tools to encourage their re-use in further applications.

D.2 Value to the JISC Community

17. Achieving these objectives will provide significant benefits to this JISC call on text mining use cases, to the e-Infrastructure community engagement programme, and to the e-Framework Programme as a whole. Specifically we will:

- work with the other JISC text-mining use case project to demonstrate the potential of text mining to advance social science research;
- help establish requirements for a generic toolkit of interoperable text-mining services which can be integrated into different research practices;
- investigate potential barriers to adoption, and establish responses such as user training and support. In pursuit of this we will work closely with the JISC-funded 'Barriers to Adoption of e-Infrastructure' project which is led by NCeSS;
- document the application of text mining tools for frame analysis as an e-Framework use case. In pursuit of this, we will work closely with the JISC-funded 'Use Cases and Service Usage Models' project in which NCeSS is a partner.

18. Working closely with the other JISC Text mining project and building on NCeSS's related research programmes, this project will also help establish the foundations for the wider adoption and sustainability of NaCTeM text-mining services.

D.3 Project Workplan

D.3.1 Requirements Gathering and Tools Customisation

19. We will build example corpora from the full-text news sources included in the digitised archives of several UK newspapers as provided by Lexis-Nexis and Thomson Gale and licensed for research use by members of the University of Manchester Library. This will allow us to conduct a comparative study on how an event is framed in different newspapers. Newspapers are widely recognised in the social science research community as an excellent source for tracking shifts in the 'mood of the day' about topics of public concern, and they attempt to influence everyday public issues through the ways in which they frame their accounts of the events of the day. They are well-suited to longitudinal analysis of shifts over time.

20. We will use a combination of methods to ensure that user requirements are systematically identified and tracked over the course of the project. These will include ethnographic studies of frame analysis methods, discussions, interviews and workshops. Ethnography is today widely used for IT systems requirements gathering.¹⁵ Its value lies in its focus on the detailed study of how work actually gets done. This will be vital if we are to understand how to embed text-mining services within a context of established routines of social research practice and resource use.

21. Based upon the understanding of the frame analysis process gained from the requirements investigation, we will customise the ASSERT tools for assisting in the tasks of text summarisation and classification and develop a simple web-based portal interface. Among the issues we will investigate are: the kinds of options users will need to be able to set to define the behaviour of the ASSERT tools; user requirements for interacting with and controlling or 'steering' the analysis process, including setting check points and backtracking the analysis to previously set checkpoints; the visualization of intermediate and final results; comparing outcomes of, for example, different coding schemes; and capturing a record of the process so that it may be re-run or shared with other researchers.

¹⁵ Anderson, R. (1994). Representations and requirements: The value of ethnography in systems design. *Human-Computer Interaction* 9, p. 151-182.

22. The project will follow an iterative, user-centred methodology of requirements gathering, rapid prototyping, evaluation and refinement to ensure that we are able to track how requirements evolve as users begin to apply new tools in their work.¹⁶ To ensure this tight coupling between requirements gathering, development and evaluation, we will co-locate the frame analysis methods, user requirements and text mining RAs, thereby ensuring that user participation is seamlessly incorporated throughout the whole development process.

D.3.2 Final Evaluation

23. A final evaluation will be conducted in order to document in an objective way the impact of text mining on frame analysis. The text-mining research community has established several performance evaluation benchmarks and metrics. While these are essential to prove the performance of algorithms and systems, they are not enough to answer the question that this project poses: do the text mining tools improve the frame analysis process and how can it be further refined? Text mining performance benchmarks, moreover, are statistical and indicative, and need careful interpretation to be useful.¹⁷ Furthermore, frame analysis is an iterative and interactive process where results from each stage are assessed by the researcher and used to decide on the next step. It is this interaction between tool and user which the evaluation process for the text-mining tools needs to be able to capture: in other words, evaluation must be done *in context*. Effective tools must not only meet performance requirements but also mesh with existing and developing social research work practices and processes. Going from simple tasks (where objective metrics can be easily defined) to more difficult and realistic tasks is a challenge.¹⁸ Evaluation procedures need to be both repeatable, so that performance and usability improvements may be quantified, yet also capable of exposing for analysis the impact of the text mining tools on the frame analysis process.

24. We will define an evaluation framework which addresses both the objective and subjective perspectives and incorporates both quantitative and qualitative approaches. Evaluation will be performed in two ways: (i) on the individual text-mining modules and (ii) on the overall set of tools, in each case to examine whether they usefully contribute to and support frame analysis.

25. In conjunction with CRESC, we will define objective metrics for the component parts of the text-mining tools (text clustering, classification and summarisation modules) and use these to evaluate them individually. We will also design a set of reference tasks which are representative of frame analysis as an iterative, interactive process and which will enable us to evaluate the system in context. Parameters for inclusion in the set of reference tasks include: frequency, criticality, representativeness, etc. Using the evaluation framework, we will design a set of procedures for collecting objective measures of performance together with subjective measures, and investigate the impact of the text-mining tools on frame analysis processes. These procedures will include benchmark tests, prototyping workshops, experiments and observational studies of frame analysis.

26. Researchers at CRESC continue to conduct research in media studies to see how emerging events are portrayed in national and international newspapers. Both 'prospective' and 'retrospective' evaluations can therefore be conducted. Retrospective evaluations will take studies already coded and analysed using frame analysis and assess the extent to which using text mining might have changed – and specifically improved -- the process and outcomes. These data will be critical components in the early testing and development phases. Prospective evaluations will be conducted on studies that are on-going at CRESC during the project. These evaluations will comprise of two parallel processes: one will be a frame analysis conducted in the usual way and the other will be an analysis, answering the same question, which uses text mining. These evaluations will examine:

16 Hartswood, M., Procter, R., Rouncefield, M., Slack, R. and Voss, A. Co-realisation: Evolving IT Artefacts by Design. In Ackerman, M., Erickson, T. and Halverson, C. (Eds.) *Evolving Information Artefacts*, Springer, 2007.

17 Underwood, N. and Lisowska, A. (2006). The Evolution of an Evaluation Framework for a Text Mining System. In *Proceedings of 5th International Conference on Language Resources and Evaluation*, Genoa, Italy., pp 2479-2484.

18 Whittaker, S. Terveen, L. Nardi, B. A. (2000). Let's Stop Pushing the Envelope and Start Addressing It: A Reference Task Agenda for HCI. *Human Computer Interaction*, vol 15(2/3), pp 75-106.

- the potential of text mining to act as a second coder (for quality control) for screening studies; this will assess both the reliability of the text mining tools and the type and quantity of information required for the researcher to verify that the text mining tools have performed as expected.
- quantitative issues, such as which studies were included/excluded under the two methods, and why.
- qualitative issues, such as interaction between researcher and the tools: how much prior knowledge about the internal processes of text mining does the researcher need in order to use it effectively? How long does learning to use the text mining tools effectively take? What kind of documentation is needed and what aspects of the user interface help/hinder this process?
- the extent to which text mining changes the frame analysis process and quality of the results. One of the most important issues to be evaluated will be how text mining innovates and advances thinking of what constitutes a frame analysis. Current work at CRESC often begins an analysis by 'mapping' the scope of research in a given area and uses the results to start analysing the media content. Text mining may prove especially useful in this early stage, enabling researchers to 'map' an area quickly and be more responsive to generating frames.

D.4 Collaborator Roles

27. The National Centre for e-Social Science (NCeSS), working in collaboration with the ESRC Centre for Research on Socio-Cultural Change (CRESC)¹⁹, will be responsible for the conduct of requirements gathering, prototyping, testing and evaluation. NCeSS will also be responsible for overall project management. NaCTeM, working in collaboration with NCeSS and leveraging the ASSERT project²⁰, will be responsible for customising the ASSERT text mining tools. Peter Golding, the leading media researcher in the UK, will be responsible for outreach and leading the engagement with the wider social scientific research community. We will host one workshop to present the use of text mining in frame analysis and to obtain feedback from the wider user community on our findings.

D.5 Project Management

28. The Principal Investigator will oversee the project and be responsible for overall co-ordination. The co-investigators will be responsible for management of the particular research activities associated with their organisation and act as the organisation's primary point of contact. To facilitate the collaboration between two JISC text mining projects, we will invite the PI from the second project to be involved in our project managerial activities. The investigators collectively will make up the Management Committee which will formally meet quarterly (by Access Grid and face-to-face). Its main activities will focus on progress assessment and early detection of problems, proposing and approving changes to project plans in response to unexpected problems or conflicting situations, and co-ordination of the project across the participating institutions.

29. Regular reviews will be used to track progress, mitigate risk and amend the research plan when needed. The general structure of the review process will be:

- The circulation of interim reports and end of year reports.
- The presentation of the research and findings at the workshop.
- Review of progress by the Management Committee.
- Report on progress by the Principal Investigator providing a formal review of progress and recommendations made by the Management Committee.

D.6 Programme of Work

30. The schedule of work packages, activities, deliverables and staff effort is laid out in the table below with the responsible partner defined in table 1.

WP	Months	Description	Deliverables	Effort person months
-----------	---------------	--------------------	---------------------	-----------------------------

¹⁹ <http://www.cresc.ac.uk/>

²⁰ <http://caqdas.surrey.ac.uk/>

WP1 NCeSS	1	Meeting with CRESC to refine scope of requirements investigation and arrange dates for fieldwork. Corpus collection.	D1.1 (M1): Plan of study and agenda for implementation, core set of study questions. D1.2 (M1): Training and test corpus.	NCeSS 0.5 CRESC 0.5 NaCTeM 0.5
WP2 NCeSS	1	An initial period of research into existing literature on frame analysis methods, tools and technologies, usability issues; surveys of relevant technologies and standards.	D2 (M1): Report providing context and related social and technical experiences of frame analysis.	NCeSS 0.5 CRESC 0.5
WP3 NCeSS	2-3	Ethnographic study of frame analysis practices, interviews and discussions.	D3.1 (M3): Report on frame analysis practices and requirements for first text mining prototype.	NCeSS 0.5 CRESC 1
WP4 NCeSS	4	Initial requirements specification. Evaluation framework.	D4.1 (M4): Text mining tools specification. D4.2 (M4): Evaluation framework, tasks, metrics and procedures definition.	NCeSS 0.5 NaCTeM 0.5
WP5 NaCTeM	5-10	Iterative text classification and summarisation modules customisation, testing and refinement.	D5.1 (M10): Text classification module. D5.2 (M10): Text summarisation module.	NaCTeM 3 CRESC 2 NCeSS 1
WP6 NaCTeM	11-12	Design, build and evaluate web-based portal user interface. Integration of tools and user interface modules.	D6.1 (M12): User interface module. D6.2 (M12): First prototype.	NaCTeM 1 NCeSS 0.5 CRESC 1
WP7 NaCTeM	12-14	Benchmarking, experiments and observational studies of use of text mining as defined by evaluation framework.	D7.1 (M14): Evaluation report: performance and usability issues for text mining services in frame analysis. D7.2 (M14): Revised text mining tools specification. D7.3 (M14): Metadata schema for textual coding.	NCeSS 1 CRESC 2 NaCTeM 2
WP8 NaCTeM	15-16	Final prototype development and evaluation.	D8.1 (M16): QDA Text mining v.1.0 built and released. D8.2 (M16): Final report on requirements, performance. D8.3 (M16): Evaluation framework.	NCeSS 1 NaCTeM 1 CRESC 1
WP9 Peter Golding	17	Workshop at which project results will be presented, other applications of text mining techniques for the social sciences explored. Workshop proceedings will be published on NaCTeM and NCeSS websites.	D9.1 (M17): Workshop proceedings published. D9.2 (M17): Report on wider applicability of text mining. D9.3 (M18): Training and support materials. D9.4 (M18): e-Framework Use case. D9.5 (M18): Report on barriers to adoption of text mining services.	NCeSS 0.5 NaCTeM 1 CRESC 1 UKDA 1
WP10 NCeSS	18	Final project report.	D10 (M18): Final project report.	PI and Co-Is.
WP11 NCeSS	1-18	Project management.	D11 (M6, 12, 18): Interim and annual reports.	NCeSS 1

D.7 Deliverables

31. The deliverables of this project are (see table above):

D7.1 Software Outputs

- a frame analysis demonstrator based on a customised version of current ASSERT tools (D8.1);

D7.2 Non-Software Outputs

- a case study demonstrating use of ASSERT text mining tools in social sciences (D2, D3.1, D7.1, D8.2, D9.6)
- Evaluation framework (D8.3);
- e-Framework use case documenting processes and workflows (D3.1, D9.4);
- Training materials (D9.3);
- Report on barriers to adoption and strategies for tackling them (D9.5);
- Final report summarising major findings, lessons learnt and the impact of this project's approach to facilitating effective frame analysis (D10).

D.8 Risk Management

Risk	Probability (1-5)	Severity (1-5)	Score (PXS)	Action to Prevent/Manage Risk
Staffing	2	4	8	People might leave. The collaborating institutions work hard to provide convivial working arrangements and clear career development paths for their staff.
Organisational	4	1	4	There are difficulties involved in collaborating between more than one institution. However, all institutions are based at the University of Manchester. And project teams have experience in cross-discipline work.
Technical	3	3	9	The technologies involved are challenging. However, the sites involved are experts and well known to appropriate user groups if external assistance is required.
External suppliers	1	1	1	No external suppliers involved.
Legal	2	1	2	The software will be released under open source software licences and we have obtained informed consent of using the confidential interview data. So the legal risk is reduced largely.
User stakeholders	2	4	8	Users are so highly integrated in this project that significant problems would occur were they not to be fully engaged. We have mitigated this risk through embedding the observer/evaluator throughout the development and implementation process, which improves the interactions with the users. Additionally, the co-Investigator Colette Fagan and Louise Corti have worked with many sociologists and this would provide good route for dissemination and outreach.

D.9 Intellectual Property Rights (IPRs) Issues

32. NaCTeM text mining services will be made freely available for use. Other software developed in this project will be available on an open source basis, licensed for free non-commercial use and development and will be available to the UK HE and FE community in perpetuity. We will publish our data coding metadata schema and through the UKDA, work with DDI to encourage open standards.

D.10 Sustainability

33. This project regards sustainability as a priority if full value is to be achieved. We will

address this issue in the following ways. NCeSS will continue development of the text mining portal and will host it on its ERSC-funded e-Infrastructure it is building for the social science community.

34. In the future, we will integrate these text mining tools with other e-Research tools for linking, processing, managing and sharing multiple forms of social scientific data. Our commitment to open standards will help in the achievement of sustainability by encouraging interoperability and tool integration.

E Engagement with the Community

E.1 Dissemination

35 We will utilise NCeSS's collaboration with NCRM nodes which focus on the innovation, integration and use of qualitative research methods within the social sciences²¹ to assist with community engagement. We will also exploit NCeSS's extensive programme of outreach to the social science research community, including participation in events such as the Research Methods' Programme Summer Festival.

36. We will to host a user workshop to demonstrate the text mining tools to the wider social science research community and gather requirements for a future service roll-out by NCeSS and NaCTeM. This work will be guided and orchestrated by Prof. Peter Golding and the Communications Research Centre at Loughborough University.

37. All published output of this project, including software, software documentation, research papers and conference proceedings arising from this work will be deposited at the JISC repository 'The Depot', and made accessible on the NCeSS and NaCTeM websites.

E.2 Contribution to JISC Programmes

38 We have allocated at least 10 person-days per year and related expenses to engage in JISC Capital Programme-level activities as part of our project plan, plus two person days per year to contribute to the e-Framework programme. We will attend programme meetings and relevant special interest groups.

F Budget

Directly Incurred Staff	April 07 - March 08	April 08 - March 09	Total £
Developer, ██████████ (NaCTeM)	██████	██████	██████
Yuwei Lin, ██████████ (NCeSS)	██████	██████	██████
Driver Project 2, ██████████ (NCeSS)	██████	██████	██████
Project Manager, ██████████ (NCeSS)	██████	██████	██████
Administrator, ██████████ (NCeSS)	██████	██████	██████
Total Directly Incurred Staff (A)	27628	60178	87806
Non-Staff	April 07 - March 08	April 08 - March 09	Total £
Travel and expenses	1425	2850	4275
Hardware/software (Laptops x 2)	1500		1500
Dissemination	0	2000	2000
Evaluation			
Consultancy fee: ██████████	██████	██████	██████
Other			
Total Directly Incurred Non-Staff (B)	6425	11350	17775
Directly Incurred Total (A+B=C)	34053	71528	105581
Directly Allocated	April 07 - March 08	April 08 - March 09	Total £

21 <http://www.cardiff.ac.uk/socsi/qualiti/>

Staff			
- Peter Halfpenny (5%)			
- Farida Vis (OU) (5%)			
Estates	4909	10009	14918
Other			
Directly Allocated Total (D)	8334	17467	25801
Indirect Costs (E)	30806	62802	93608
Total Project Cost (C+D+E)	73193	151797	224990
Amount Requested from JISC	58554	121438	179992
Institutional Contributions	14639	30359	44998
Percentage Contributions over the life of the project	JISC 80%	Partners 20%	Total 100%

G Previous Experience of the Project Team

39. The investigators bring together a body of experience to manage the research and its outputs.

40. **Prof Peter Halfpenny** has been Director of the Centre for Applied Social Research for over 20 years, Professor of Sociology for 15 years and Executive Director of the ESRC-funded National Centre for e-Social Science (NCeSS) since its inception in 2004, all at the University of Manchester, where he has been since 1972. His role in NCeSS is overall strategic management of the Co-ordinating Hub, seven Research Nodes and 12 Small Grant Projects. He is responsible to the ESRC for the success of the Centre's programme of research, outreach and capacity-building, as well as advising the ESRC on its e-Science strategy. His own research interests are in the integration of Grid tools and services into a comprehensive support environment for social science researcher practitioners, and in the investigation of the adoption and adaptation of computational support tools across the social research community. He was, until recently, Director of the University of Manchester's MA programme in Applied Social Research, and he has over 20 years' experience of teaching social research methods at undergraduate and postgraduate levels.

41. **Dr Farida Vis** is a Research Fellow at the ESRC Centre for Research on Socio-Cultural Change (CRESC), which is an interdisciplinary collaboration based at the Open University and the University of Manchester. Her research falls within CRESC's 'Media and Economic and Social Change' theme and it concentrates on issues of representation of war and conflict, with particular emphasis on the ways in which different victims and perpetrators have been represented with the UK and US print media. Her work has a strong methodological focus, in which frame analysis plays a key part. She has investigated how frame analysis might be standardised and applied more coherently and transparently.

42. **Peter Golding** is Professor of Sociology and Pro-Vice Chancellor (Research) at Loughborough University, and co-director of its Communications Research Centre. He co-edits the European Journal of Communication and is chair of the European Sociological Association Media Research Network. He was founding chair of the Standing Conference on Cultural, Communications, and Media Studies from 1992 until its transformation into the Media, Cultural and Communication Studies Association (MeCCSA), of which he is Honorary Secretary. He was a member of the RAE panel for Communication, Cultural and Media Studies in 1996 and 2001, and will be chair of this panel in RAE2008. In 2002 and 2005 he was Vice-Chair of the ESRC Sociology Panel for the Postgraduate Recognition Exercise. He has served as a member of the HEFCE Media Studies Advisory Group and is currently on the Advisory Panel for the ESRC e-Society Programme. In his most recent project, he conducted a study of new methods in medial content analysis, as part of the ESRC's Research Methods Programme.

JISC
Northaven House
Coldharbour Lane
Bristol
BS16 1QD

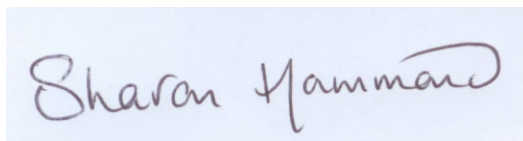
21th June 2007.

**Re: JISC Capital Programme: E –Research: e-Infrastructure
Call 1 – Automatic Summarisation for Systematic Reviews using Text Mining**

The proposed project ‘Using Text Mining for Frame Analysis of Media Content’ is in response to the Call for Projects in JISC circular 01/07 Capital Programme.

This is to confirm that if the proposal is successful, the University of Manchester will be pleased to administer and accommodate this project.

Yours sincerely,

A handwritten signature in blue ink on a light blue background. The signature reads "Sharon Hammond" in a cursive script.

Sharon Hammond
Research Development Manager
School of Social Sciences
University of Manchester.

21st June 2007

Professor Peter Halfpenny
National Centre for e-Social Science
University of Manchester
Dover Street
Manchester
M13 9PL

Dear Peter

I am pleased to support your proposal on text mining. It offers the potential for exciting methodological advances which would complement the work of NCRM. Moreover, I should welcome the opportunities for NCESS-NCRM collaboration that it could promote, most obviously via the involvement of the NCRM node on Methods of Research Synthesis, but also through interaction and synergy in broader settings.

Yours sincerely



Professor Chris Skinner
Director, ESRC National Centre for Research Methods