



Project Document Cover Sheet

Project Information			
Project Acronym	Gold Dust		
Project Title	Gold Dust		
Start Date	2 January 2008	End Date	31 March 2009
Lead Institution	University of Hull		
Project Director	Sue Geale (s.e.geale@hull.ac.uk)		
Project Manager & contact details	(Interim) Dr Caroline Ingram caroline@csiconsultancy.co.uk +44 (0)7979 596586		
Partner Institutions	Heriot-Watt University, Bath University, NaCTeM, Strathclyde University, Cambridge Information Group, University of Bolton, University of Liverpool, University of Illinois at Urbana-Champaign, Institution of Civil Engineers, Intute SET, Cranfield University		
Project Web URL	http://www.hull.ac.uk/golddust/		
Programme Name (and number)	Users and Innovation: Next Generation Technologies and Practice Phase 2 (b) Large-scale institutional demonstrators		
Programme Manager	Lawrie Phipps		

Document Name			
Document Title	Project Plan		
Reporting Period			
Author(s) & project role	Caroline Ingram (interim project manager)		
Date	December 2008	Filename	0712 Gold Dust project plan
URL	<i>if document is posted on project web site</i>		
Access	<input checked="" type="checkbox"/> X Project and JISC internal	<input type="checkbox"/> General dissemination	

Document History		
Version	Date	Comments
1.0	29/12/08	First draft for comment and amendment by project partners
1.4	06/02/08	Final Draft Submitted to JISC



Gold Dust Project Plan

Overview of Project

1. Background

1.1 Gold Dust is a project intending to investigate and develop innovatory solutions to the problem of information overload within the academic environment.

1.2 Gold Dust will build upon some by-products of the JISC Users & Innovation funded ticTOCs project. This is in response to analyses of various reports, user requirements and the outcomes of several user engagement activities across a series of JISC programmes and related initiatives. Gold Dust aims to produce and test a prototype and consequently a demonstrator for the delivery of highly relevant personalised current awareness content of a variety of kinds to academics. This should ultimately be possible without the need for any input by those academics in the personalisation process. It will test the delivery of this content from within selected JISC and non-JISC presentation services (i.e. flexible distribution of matching content), including a university institutional setting, a desktop tool, and selected web-based services including a commercial publishing service.

1.3 The project will develop Personal Interest Profiles (PIPs) from existing data. It will exploit the potential of RSS (Really Simple Syndication) and will aggregate content from numerous sources. It will then incorporate text mining techniques and terminological searching aids in a filtering process between the PIPs and the aggregated content. It will use the complex information landscape subject area of engineering as its test-bed, and a controlled group of fifty academic testers as its main user community, with additional engagement with the U&I Community.

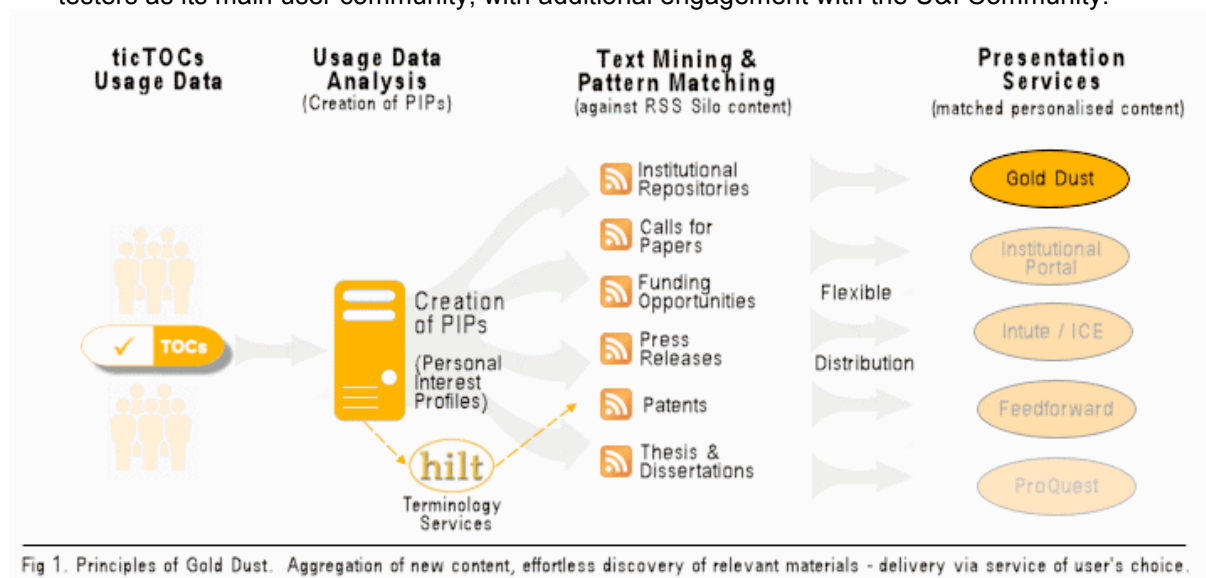


Fig 1. Principles of Gold Dust. Aggregation of new content, effortless discovery of relevant materials - delivery via service of user's choice.

In the above diagram, from the left: A test group of people are using ticTOCs in order to keep abreast of journal Tables of Contents. Their use of ticTOCs produces very interesting usage logs containing metadata that can be analysed in order to produce 'Personal Interest Profiles'. Text mining techniques are then used to match these 'Personal Interest Profiles' with content found from numerous selected, potentially relevant, RSS feeds. The results are then delivered in a flexible way at a place of the user's choice.

"Where is the data we have lost in information management, and the knowledge we have forsaken thereby"
Lorcan Dempsey (2006)¹

"The task is not to design information-distributing systems but intelligent information-filtering systems"
Herbert A Simon (1986)²

¹ Lorcan Dempsey, Emergent knowledge and intentional data, Lorcan Dempsey's weblog, December 31, 2006, <http://orweblog.oclc.org/archives/001236.html>

"If we thought that we were too busy at work before online services became available, we must be drowning today" Steven M Cohen (2003)³

- 1.4 Information overload is becoming an increasingly important issue in academia. As the online information environment grows exponentially, many academics struggle to keep up-to-date with new content relevant to their interests, both in research and learning and teaching. Current awareness is central to this process, however many existing current awareness services, for example Current Contents Connect⁴ and ACM TechNews⁵, offer relatively unfiltered content. Other services, such as PubCrawler⁶ and search alert facilities offered by various services (e.g. Engineering Village 2⁷ and Technorati Watchlist⁸) are dependent on concise, manually created and mostly static search profiles. These rarely if ever deliver personalised content which is both broad enough to reflect the real interests of academics, yet sufficiently focussed to satisfy their often very specific, and changing, information requirements. In fact, some may actually contribute to increased overload by producing long lists of potentially relevant content which require further time-consuming manual scanning and sifting, in order to separate the 'wheat' from the 'chaff'.
- 1.5 The lack of suitable current awareness solutions was recognised in the findings of the RIN study *Researchers and discovery services: Behaviour, perceptions and needs*, which reported: "A significant minority of researchers said that they avoid push services because their email is already too extensive, or that they had subscribed to alerts or listservs in the past but had cancelled them as they felt overwhelmed or the services proved of little value."⁹
- 1.6 While RSS can be seen as being a part of the potential solution to information overload, it is also part of the problem. On the one hand, RSS is increasingly enabling the delivery of current content of various kinds from various places direct to the desktop, and its potential for current awareness has already been recognised¹⁰. However, RSS is essentially publisher/content provider-driven push technology, and often results in relatively indiscriminate content which requires further subsequent human filtering. Existing RSS aggregators tend only to divert a 'river of feeds' to the end-user. Brian Kelly, in UK Web Focus has pointed out that: "With a growth in the numbers of blogs and a growth in services which output RSS, there will be a need for clever tools to help find information which is relevant to the end user."¹¹ Services such as mySyndicaat¹² and Newstex BlogAlerts¹³ have recently emerged. They aggregate, filter and deliver content from news RSS feeds. However, such services depend on 'non-smart' user-defined filtering rules, and are less sophisticated and more limited than the solutions proposed by Gold Dust. Gold Dust will develop clever, evolving, personalisation tools to filter new content contained in RSS feeds. Importantly, it will do so with little or no effort on the part of the user. It will aggregate relevant content of various kinds of specific interest to academics, and not just from news sources. In addition, Gold Dust will test the flexible delivery of results at presentation services of the users' choice. Gold Dust will aim to demonstrate the delivery of 'gold quality' information to academics.
- 1.7 The potential for better personalisation services within academia has been investigated by the *Personalisation in presentation services report*¹⁴, commissioned by JISC. The report recommended "...small pieces of work looking at user requirements and exploring innovative and tightly defined uses of personalisation approaches". The report identified "...several impediments

² Herbert A Simon "The impact of communication on organizations" in R. Wolff (ed) Organizing industrial development, 1986.

³ Steven M Cohen, Keeping Current. ALA, 2003. p. 6

⁴ <http://scientific.thomson.com/products/ccc/>

⁵ <http://technews.acm.org/>

⁶ <http://pubcrawler.gen.tcd.ie/>

⁷ Personalized RSS feeds available from Elsevier's Engineering Village 2, press release,

<http://www.reedelsevier.com/index.cfm?articleid=1426>

⁸ <http://technorati.com/watchlist/>

⁹ Researchers and discovery services: Behaviour, perceptions and needs, A study commissioned by the Research Information Network, November 2006, <http://www.rin.ac.uk/files/Report%20-%20final.pdf>

¹⁰ Jay Bhatt, Using RSS to increase user awareness of e-resources in academic libraries,

<http://www.higheredblogcon.com/index.php/using-rss-to-increase-user-awareness-of-e-resources-in-academic-libraries/>

¹¹ Brian Kelly, UK Web Focus <http://ukwebfocus.wordpress.com/2007/01/18/mybloglog/#comments>

¹² <http://www.mysyndicaat.com/>

¹³ http://www.newstex.com/company/press//Newstex_MuseGlobal_BlogAlerts.pdf

¹⁴ Nicky Ferguson, Seb Schmöller, and Neil Smith, Personalisation in presentation services, 2004, A report commissioned by JISC, <http://www.therightplace.plus.com/jp/>

to using personalisation with uncontrolled data, including immature technology and lack of metadata” and concluded that “...personalisation is effective and feasible in situations where data is controlled and where there is a clear rationale or business case”. By concentrating on logging and tracking data relating to individuals’ personal interests, produced via normal use of the JISC-funded ticTOCs project, and matching it to controlled metadata produced from selected and relevant RSS feeds of various kinds, Gold Dust will directly address the major impediments to successful personalisation identified by the report – those of uncontrolled data, lack of metadata and immature technology.

- 1.8 The Final Report of the Research Support Libraries Group (RSLG) recognises the importance of being able to sift suitable scholarly content in alerting services: “...a number of researchers...have indicated that they found particular difficulty in dealing with the sheer volume of references that conventional internet search engines (and, for that matter, subject-related “alerting services” where these are available) can provide. This suggests that there would be strong support for the development of tools to help them to sift these references, with emphasis...on quickly sorting out items produced and published to conventional research standards (including peer-reviewed material) from those that are not.”¹⁵
- 1.9 The *Personalisation in presentation services report* also stated that: “Within the JISC IE there is a potential need for personalisation to flow backwards into the content sources as well as across different services within the presentation layer. Such sharing may require a trusted intermediary to strip out personal information and possibly restore it as information comes back from the content provider to the presentation service to the end user”¹⁶. In addition, the OSI E-Infrastructure Strategy Report advised that: “Service providers need to understand in greater detail the basis for researchers’ current search and navigation-related behaviour, and to determine the extent to which the historically fragmented nature of information and data provision inhibits optimum discovery behaviour.”¹⁷ By working with various JISC IE presentation services and other content/presentation providers, by delivering content from various sources, and through testing the flexible delivery of results, Gold Dust will in effect both act as a trusted intermediary, and also enable aggregated content to flow backwards into services.
- 1.10 The proposed Gold Dust work will have potential benefits for academics and researchers, service providers and discovery services, and also content providers (including Institutional Repositories) and publishers: “Publishers have learned that the key to making a success of any electronic product is simple – make it easier for someone to do something they have to do anyway.”¹⁸ Though Gold Dust will use the specific subject area of engineering as a test-bed, its demonstrator and findings should be extendable to other subjects.
- 1.11 In summary, Gold Dust will provide potential solutions to issues identified in JISC and other reports, and several user engagement activities, and provide much value to the JISC community. Gold Dust will develop innovative, next generation procedures to improve the quality of users’ interactions with electronic information; build upon outcomes of an existing Users & Innovation programme project (ticTOCs) plus several other projects; investigate and develop personalisation techniques; encourage join-up of content from several JISC programme areas and services including repositories; deliver potential solutions in response to expressed needs; and benefit from expert input from the U&I Community.

2. Aims and Objectives

- 2.1 Gold Dust aims to develop potential solutions to deliver the right information, at the right time, to the right people, in the right way and in the right place. We aim to find practical solutions to identify highly relevant (personalised) items of interest from amongst a mass of potentially relevant current awareness information which is being generated (via RSS) by numerous content

¹⁵ Research Support Libraries Group (RSLG) Final Report, 2003 <http://www.rslg.ac.uk/final/final.pdf>

¹⁶ Op Cit, p. 34

¹⁷ OSI E-Infrastructure Strategy: Report of the Working Group on Search and Navigation, March 2006 <http://www.nesc.ac.uk/documents/OSI/search.pdf>

¹⁸ Hunter, K, Virkler, S, and Sidi, R “Disruptive technologies: taking STM publishing into the next era” *Serials* 20(1), March 2007, p.52

providers (including JISC services and projects, but also others), and then delivering it as required, to academics without requiring their input in the process.

- 2.2 The objective will be to produce and test the delivery of highly relevant, personalised current awareness content to academics, from within various presentation services, in a similar way to that described by Xin et al¹⁹. As well as offering convenience for users, this will in addition potentially improve the existing underlying information infrastructure.
- 2.3 The project will work with a selection of subject-based services and presentation services to investigate the flexible delivery of personalised content within these existing presentation services (in addition to a Gold Dust website). Some of these services generate content of their own via RSS feeds which will be included in the Gold Dust databases, and the project will investigate ways of enabling such 'home' content which matches with PIPs to be given top ranking in results delivery. Presentation services partnering with Gold Dust for this purpose include: the University of Hull institutional portal, Institution of Civil Engineers (ICE), Intute SET, Feedforward, Aerade and Cambridge Information Group (via ProQuest).
- 2.4 An additional aim will also be to test the viability of sharing user PIPs between services, to allow more personalised presentation of those services, and to investigate their potential use by ePortfolio services.
- 2.5 These aims will help to "capitalise on JISC's substantial investment in services and resources"²⁰ and provide improved navigation tools as envisaged by the OSI E-infrastructure Working Group on Search and Navigation. Gold Dust will investigate the logistics and possible benefits of sharing PIP data with a commercial publisher - the Cambridge Information Group (ProQuest, RefWorks, COS Scholar Universe, Bowker, AquaBrowser and Serials Solutions). The project intends to investigate intersections between COS Scholar Universe²¹ profile content, Scholar usage data, and PIPs.

3. Overall Approach

- 3.1 Gold Dust will collaborate with the JISC-funded ticTOCs project²². ticTOCs is a Users and Innovations project, funded from April 2007, that is developing a service to enable academics to discover and use standardised journal Table of Contents RSS feeds and their content. ticTOCs use logs will contain valuable information about the real, focussed scholarly interests of academics, and this data will be analysed in order to produce PIPs, with appropriate extra weighting given to metadata of downloaded, exported and reused items. Importantly, the PIPs will be produced without any further input from ticTOCs users other than regular use of ticTOCs. Gold Dust will also use ticTOCs' method of RSS feed aggregation to produce databases of RSS feed content for selected categories of materials as described in 3.4 below. The project will then use text-mining and pattern analysis techniques (as suggested in the OSI E-infrastructure Working Group report²³) to match individual PIPs with relevant new content collected in the databases. The potential for text-mining in this context has been recognised by Clifford Lynch: "We would also see a move beyond federation and indexing to actual text mining and analysis, to the extraction of hypotheses and correlations that would help to drive ongoing scholarly inquiry."²⁴
- 3.2 Research is required to determine the most relevant and efficient form and content of the proposed PIPs. This work will be supported by Chris McMahon of the University of Bath, and based on similar investigations carried out by Campbell *et al*²⁵, which were based on monitoring

¹⁹ Zhiyun Xin, Jizhong Zhao and Chihong Chi, Information push-delivery for user-centred and personalized service. Paper presented at the Second International Conference on Fuzzy Systems and Knowledge Delivery (FSKD), Changsha, 2005.

²⁰ Nicky Ferguson, Seb Schmoller, and Neil Smith, *op cit*, Executive Summary, p. 3.

²¹ <http://info.csa.com/communityofscholars/>

²² <http://www.tictocs.ac.uk>

²³ OSI E-Infrastructure Strategy: Report of the Working Group on Search and Navigation, March 2006
<http://www.nesc.ac.uk/documents/OSI/search.pdf> p. 13-14.

²⁴ Clifford A. Lynch, "Open Computation: Beyond Human-Reader-Centric Views of Scholarly Literatures," Open Access:Key Strategic, Technical and Economic Aspects, Neil Jacobs (Ed.), (Oxford: Chandos Publishing, 2006), pp. 185-193.

²⁵ D R Campbell, S J Culley, C A McMahon and P Coleman, A methodology for profiling computer based design activities, Paper presented at the International Conference on Engineering Design, ICED 05, Melbourne, 2005

user activity on a workstation, including interactions with office and Computer Aided Design (CAD) software packages and internet browsing habits. CiFlex Software²⁶ was an outcome of this work. Campbell *et al*, also successfully used a PIP concept to develop 'Push-based' strategies for improving the efficiency of information management in design projects. The concepts adopted will prove a good basis for the work of Gold Dust, and will help determine which information from ticTOCs' usage logs is most relevant.

- 3.3 This obtained knowledge can be used in conjunction with text-mining, pattern-recognition, and other data mining technologies to identify relevant and useful content from the databases of aggregated RSS feeds in order to provide 'push-based' information content to users. With respect to text mining, the project will build upon research already conducted elsewhere. Thelwall *et al*²⁷ identified the potential of RSS feeds as an appropriate information source from which to investigate 'broad issue scanning': "RSS feeds are potentially a non-intrusive source of high quality data about public opinion: monitoring a large number may allow quantitative methods to extract information relevant to a given need." It should be noted that, instead of 'public opinion', Gold Dust will be concerned with relevant scholarly and other content of different kinds of interest to academics. The project will benefit from input from the National Centre for Text Mining (NaCTeM) with respect to identifying existing appropriate text mining methodologies and algorithms to match the PIPs with content from the databases.
- 3.4 Categories for databases of RSS content will include, but may not be limited to: new items in Institutional Repositories and subject repositories, Calls for Papers, funding opportunity news, patents, press releases, professional society news, engineering news feeds and component announcements, teaching and learning resources, journal Tables of Contents, forthcoming conferences, theses and dissertations, and news from JISC services and projects.
- 3.5 Gold Dust will work with the HILT²⁸ project to investigate the potential benefits of using terminological searching aids to improve pattern matching between the PIPs and text-mined content from aggregated databases of RSS content. Such 'behind the scenes' matching is intended to enrich results without the need for user input or effort²⁹.
- 3.6 The project will work with the JISC-funded Feedforward³⁰ project, at the presentation level (see 2.3 above) to expose Gold Dust content as an API within the Feedforward desktop tool. It will also co-operate with Feedforward in order to share effort in RSS aggregation, and also to investigate using Feedforward in the Gold Dust content ranking process.
- 3.7 The University of Illinois at Urbana-Champaign will collaborate with the project to provide a subject based cross-search engine to supplement, at the Gold Dust presentational interface, the delivery of the envisaged personalised current awareness content. This will enable cross-searching of the unfiltered aggregated RSS content, by categories of resources (IR content, Calls, etc), and in addition it will allow search across other relevant digital repositories in engineering and technology. In this respect, it will build upon, and further develop, the work and findings of the JISC-funded PerX project, plus the TechXtra service and NSF and Mellon-funded federated search work at Urbana-Champaign³¹. In addition to providing a search service for the Gold Dust site, this will help identify sources of further RSS feeds for inclusion in the databases of RSS content, and will provide additional usage logs which may be used in PIP enhancements.
- 3.8 Gold Dust will work with RefWorks³² at the presentation stage to investigate RefGrab-It possibilities for downloading non-standard citations from results into RefWorks.

²⁶ http://people.bath.ac.uk/enpdr/Ciflex_Overview.htm

²⁷ Mike Thelwall, Rudy Prabowo, Ruth Fairclough, Are Raw RSS Feeds Suitable for Broad Issue Scanning? A Science Concern Case Study
<http://www.scit.wlv.ac.uk/~cm1993/papers/Are%20rss%20feeds%20suitable%20for%20broad%20issue%20scanning%20preprint.doc>

²⁸ <http://hilt.cdli.strath.ac.uk/>

²⁹ Tseng, C, and Ng, P "Precisiated information retrieval for RSS feeds" Information Management & Computer Security, 15(3), 2007, pp.184-200

³⁰ http://www.jisc.org.uk/whatwedo/programmes/programme_rep_pres/tools/feedforward.aspx

³¹ The current version of the Grainger Engineering Library Information Center portal is available at:

<http://search.grainger.uiuc.edu/top/>

³² <http://www.refworks.com/>

- 3.9 It should be noted that Gold Dust will not be concerned, in the initial stages at least, with collaborative filtering of content concepts, but during the second iterative process the project will investigate possibilities such as matching appropriate content found from one user's PIP to other users' PIPs with similar characteristics.
- 3.10 Gold Dust will review existing services and solutions which could have a bearing on the proposed work. These include Yahoo Pipes <http://pipes.yahoo.com/> (RSS feed aggregator and manipulator), Dapper <http://www.dapper.com/> (Access and content distribution), afeeda <http://www.afeeda.com> (aggregation), AidRSS <http://www.aiderss.com/> (RSS filtering), Particls <http://www.particls.com/> (RSS ranking), xFruits <http://www.xfruits.com/> (RSS mashup creation), Feed Digest <http://www.feeddigest.com/> (mixing and publishing feeds), Blastfeed <http://www.blastfeed.com/> (RSS filtering), Scintilla <http://scintilla.nature.com/> (aggregation and filtering), and Yahoo Panama (personalising adverts).
- 3.11 The project will last for fifteen months from the programme start date of 2nd January 2008.
- 3.12 There are two types of critical success factors, those for the project, and those which act beyond the project lifetime. For the project:
- Recruitment of 50 engineering community users;
 - Successful analysis of ticTOCs usage data to enable the creation of PIPs;
 - Successful use of text-mining and pattern analysis techniques to match PIPs with content;
 - Delivered content matches interests of community users;
 - Increased relevance of matching content during phase 2 testing;
- and beyond the project lifetime:
- Increased uptake of RSS by content providers;
 - Successful use of flexible delivery concept via presentation services, including the University of Hull institutional portal, ICE, Intute SET, Feedforward, Aerade, CIG.

4. Project Outputs

- 4.1 The main outputs will be an open source prototype, and consequently a full-scale demonstrator, which demonstrate the delivery of highly personalised current awareness information on a regular basis in a flexible way (at the right time, in the right way to the right person, in the right place).
- 4.2 The prototype will be formally tested and improved, and the resulting demonstrator will be formally tested again with the community user group in real life conditions.
- 4.3 Other deliverables will include a specification for PIPs, databases of aggregated feed content of different types, a report on potential commercial applications, and evaluation, sustainability, progress and final reports.

5. Project Outcomes

- 5.1 The project will improve the quality of users' interactions with e-systems and benefit the JISC user community by investigating and developing innovative ways to facilitate the discovery and delivery of new, highly relevant resources (including, but not restricted to, content from, and about, JISC services and projects) with ease, with potential benefits for all academics in Higher and Further Education.
- 5.2 There are also a number of outcomes of benefit to the lead and main partners, and, through dissemination of adopted practice, to the JISC community as a whole:
- 5.2.1 The Gold Dust project will allow The University of Hull to build on findings from the PORTAL and CREE projects, and take a further step towards the ability to deliver personalised and focused information through an institutional portal environment. In targeting an institutional portal we are conscious of allowing users to surface the information where they need it, and

- will also investigate the scope for delivering Gold Dust outputs through a VLE, using Sakai as our exemplar, and take on board lessons from Gold Dust in considering additional options. The project will sit alongside internal work by the University of Hull Library to develop its online presence and develop services that take the library to the users where they are online.
- 5.2.2 Gold Dust will allow Heriot-Watt University to build upon various existing and past research initiatives such as research within the School of Engineering and Physical Sciences (EPS) on shape matching and data mining, and ICBL research and development in RSS and engineering information. The university is involved and has invested in ticTOCs, TechXtra, PerX, Intute SET, Developing Personalisation in the Information Environment (DPIE), plus previous PALS Metadata and Interoperability Projects, and Gold Dust will enable the EPS and ICBL to further showcase some of its research and development strengths, with possibilities for more R&D in the future. Publications and papers arising from the project will enhance the institution's future research assessment exercise.
- 5.2.3 Cambridge Information Group: CIG includes businesses (ProQuest, RefWorks, COS Scholar Universe, Bowker, AquaBrowser and Serials Solutions) which provide services and products to the scholarly research community. From alerting services for calls for papers to dissertations RSS feeds, the strategic direction of these businesses will benefit the exploration of text mining for PIP development. CIG participation in the Gold Dust project is intended to guide the development of technology solutions for the industry that will enable cross-collaboration, sharing and delivery of content.

6. Stakeholder Analysis

6.1 The following stakeholders have been identified as having an interest in the outcomes from Gold Dust:

Stakeholder	Interest / stake	Importance
Academics & researchers	Personalised information delivered in a way that facilitates use and knowledge management, and reduces retrieval effort.	High
Publishers and content providers	An approach that allows RSS feeds to be more effectively utilised and provides better return on investment	High
Librarians	An added value means for highlighting the benefits of licensed and quality resources and a means for effectively delivering access to these	High/Medium
Institutional Repository community	An approach that develops methods to expose new content to target audience	High/Medium
Portal managers/service providers	An additional value-added service that enhances the services currently delivered and allows a greater degree of personalisation	Medium/High
JISC services and projects	An approach that allows RSS feeds to be more effectively utilised and delivered to interested users	Medium/High
Event organisers	The opportunity to effectively associate event information and content with related information from other sources and be delivered to interested users	Medium/High
Text mining community	An exemplar of how text mining can be used to target information to individuals	Medium/Low
Students	A way of filtering information to facilitate better knowledge management	Medium/Low

6.2 As well as users, project stakeholders will include those Partners and Project Associates whose content is included in the databases of RSS content, and/or whose services are included in the flexible delivery of content, and feedback from them on the evaluation process will be invited.

7. Risk Analysis

Risk	Probability (1-5)	Severity (1-5)	Score (P x S)	Action to Prevent/Manage Risk
Staffing	1	4	4	Most staff associated with this proposal are currently in post and are available from the proposed start date. A rapid recruitment process will secure a project manager.
Organisational	1	4	4	The Steering Committee is structured across several institutions/organisations. Clear project documentation will detail work packages. Lead site has project management expertise. Partner sites have considerable project development experience. Commitment from partner organisations. A partnership agreement will be secured at the start of the project.
Technical	1	2	2	The hardware to be hosted at MIMAS, and used at MIMAS and Heriot-Watt will be well managed & maintained. Data is regularly backed-up
Failure of text mining techniques	2	4	8	Project partners have experience, and can call on various connections for assistance, and re-try alternative matching techniques if necessary.
Legal	1	3	3	IPR will be respected, permission will be sought from publishers where necessary
Software failure	1	2	2	The project will use established or supported technology where possible, with on-going development commitment.
Lack of suitable content	1	2	2	Many RSS feeds are being produced by publishers and content providers, and the trend is upwards
Delays to ticTOCs development	2	3	6	ticTOCs initial prototype could be used until a ticTOCs service is available
Inadequate ticTOCs usage log data	1	2	2	A 'Paper Pilot' will explain requirements to closed user group community, including the need for their proactive use of the ticTOCs service in order to create sufficient data for Gold Dust use.
Some databases of RSS content produce no 'gold dust'	5	1	5	Variety of content types ensures relevant content from other databases. Pattern matching techniques may well need to vary, depending on which database is being interrogated, though some may fail to produce relevant content.

8. Standards

- 8.1 The Gold Dust project will make use of open standards wherever possible, focusing around the use of both versions of RSS and ATOM feeds as the source content from which it will develop the PIPs through text mining and pattern matching. It is anticipated that web services will be used where feasible. TicTOCs is already collecting Atom feeds.
- 8.2 Documentation of the technology will be a key part of ensuring the sustainability of the Gold Dust outputs. This will include the production of one or more SUMs, at a level of granularity to be agreed with the JISC, using recommended e-Framework templates.

Name of standard or specification	Version	Notes
RSS	1.0, 2.0	
ATOM	1.0	
OPML	1.0	for aggregating big batches of feeds
SRU	1.2	for a light REST-based web services support

9. Technical Development

- 9.1 Major technical components of the project are the "information push-based" algorithms and systems, and the text mining and pattern matching techniques. These techniques and algorithms will be developed using the Java language. The system applications for producing the categories of databases of RSS feeds, for tracking their usage as well as for their consume from the web will be implemented using mature Open Source technology such as the Lucene search engine and the MySQL Database as well as reusable software components produced by the current ticTOCs Project and the previous PerX project.
- 9.2 Java-based technology and Java Servlets will be given priority to increase the likelihood of software compatibility with APIs and software applications.
- 9.3 The technical development will consist of two phases and will use an iterative approach to produce the software deliverables. Phase 1 will see the creation of PIPs, and the use of pattern matching and text mining to match them with content in categories of RSS feeds. Phase 2 will see improved methods to do the same, and the addition of HILT integration.
- 9.4 See Appendix C for a technical development timeline (separate MS Excel document).

10. Intellectual Property Rights

- 10.1 Any software created in the course of the Gold Dust project will be Open Source: licensing and distribution will be guided by advice from OSS Watch. Publishers and content providers own the copyright on the RSS feeds that Gold Dust will use. Most publishers state that their feeds may be re-used, often specifying that the text of feeds must not be altered and must include the publisher's copyright statement. The Gold Dust project will ensure that IPR are respected and, where necessary, will seek permission from the publishers.
- 10.2 Commercially sensitive information about Cambridge Information Group and its products arising from the project will not be made public.

Project Resources

11. Project Partners

Partner	Contact	Role
University of Hull Library (lead)	Sue Geale	Project Director; User

		testing, Flexible delivery
Heriot-Watt University – EPS, ICBL, and Library	Heather Rea	Research into profiling, technical development
Bath University	Chris McMahon	Pattern matching
NaCTeM	Dr Sophia Ananiadou	Text mining
Strathclyde University	Dennis Nicholson	HILT integration
Cambridge Information Group	Helle Lauridsen	Technical development guidance
University of Bolton	Scott Wilson	Feed Forward integration
Liverpool University	Terry Bucknell	Technical development
University of Illinois at Urbana-Champaign	William Mischo	Technical development
Institution of Civil Engineers	Mike Chrimes	Flexible delivery
Intute SET	John Blunden-Ellis	Flexible delivery
Cranfield University	John Harrington	User community group

11.1 We will aim to draft the consortium agreement in February to send to the project partners for agreement and signing. As soon as all the signatures are received we will forward a copy to the programme manager.

11.2 Experience of the main project team:

11.2.1 Sue Geale (Project Director), Hull, is a senior manager within Library Services with many years experience of financial, project and staff management. She is a member of a wide range of University committees, including working groups developing an institutional repository for the University and the move to submission and retention of e-theses. She is also Hull's primary contact for their participation in the JISC UK LOCKSS Pilot.

11.2.2 Dr. Santiago Chumbe (Technical Development), Heriot-Watt, has been a research associate at the ICBL since 2000. He is the ticTOCs Technical Manager and works on the technical sides of VLEs, federated searching and e-journals. He is the technical manager of TechXtra, which includes two RSS aggregation services.

11.2.3 Dr Heather Rea (Research Support and Co-ordination), Heriot-Watt, has worked as a Research Associate on four EPSRC funded projects dealing with adaptive interfaces, 3D Computer Aided Design (CAD) model shape matching and data mining, and engineering knowledge capture and management.

11.2.4 Roddy MacLeod. (Subject advice, support and liason), Heriot-Watt, is Senior Subject Librarian, TechXtra Manager, previously PerX Manager and EEVL Manager, editor of *Information sources in engineering*, KG Saur, 2006, and author of numerous articles on engineering information.

11.2.5 Chris Awre (Adviser) was Project Manager for the CREE project and is currently managing the CREE Extension project, both of which were funded by JISC. He chairs the University of Hull Library Systems Strategy Working Group and is also leading the implementation of the University of Hull institutional repository.

11.2.6 Steve Hitchcock (Adviser) is currently Project Manager for the JISC-funded PRESERV2 project, and also manages the EPrints Community Programme. He has been active in the open access community and involved with a number of JISC projects for 10 years.

12. Project Management

12.1 The project will be managed by the University of Hull, and its development will be overseen by a Steering Committee consisting of representatives from the project partners, plus two advisers (Steve Hitchcock, University of Southampton, and Chris Awre, University of Hull). The role of the Steering Committee will be to oversee the project, endorse any proposed major changes to the project plan (subject to approval by JISC U&I), advise the project manager and represent the interests of the partners and project associates. The Committee is expected to meet physically once, and to conduct virtual meetings as required.

12.2 Initially the project will be managed, on a part time basis, by Dr Caroline Ingram (CSI Consultancy Ltd), an independent project consultant, with 5 years independent experience,

following roles as programme manager for JISC, ESRC and EPSRC. The project is advertising for a full time project manager, it is anticipated that this post will be filled by April 2008 as a one year contract.

12.3 The project consortium consists of Project Partners, and Project Associates. Within the Partners are institutions and organisations with extensive experience of R&D, RSS, service presentation and delivery, text mining techniques, content delivery, dissemination, institutional portals, and working with the community via focus groups, etc.

12.4 Project Associates have expressed their support for the project and will make their content available (note that databases of RSS content will extend to many other sources as well); they are:

- Morgan & Claypool. Contact: Mike Morgan.
- Oxford University Press (Oxford Journals). Contact: Kirsty Luff
- Jorum. Contact: Christine Rees
- SPIE. Contact: Jackie Sentenne Pettit
- Inderscience Publishers. Contact: Mohammed Dorgham
- Pro-Talk. Contact: Chris Rand

13. Programme Support

13.1 Ongoing advice on UIDM as we proceed highlighting potential benefits and paths we might consider

14. Budget

14.1 The project budget for Gold Dust is £258,081. The contribution requested from JISC is £197,312. The remaining costs of £60,769 will be met by the consortium partners.

14.2 The budget is attached at Appendix A. There have been no changes since the proposal was submitted.

Detailed Project Planning

15. Workpackages

15.1 The project will run in two phases, the first from January 2008 – September 2008, the second from October 2008 – March 2009.

15.2 Workpackage detail is attached at Appendix B as requested

16. Evaluation Plan

Timing	Factor to Evaluate	Questions to Address	Method(s)	Measure of Success
Easter 2008	Progress and structure of technical and research work	Does the project have the required knowledge to proceed to technical development?	Internal review and meeting	Sign off first three months of the project and proceed to technical development
After Phase one testing	Project progress	To what extent has relevant content been identified and delivered to the target group?	User testing	Constructive feedback from users on future steps to take
After Phase two	Project progress	To what extent has the relevance of content identified and delivered	User testing and internal review	User feedback and internal acceptance of project work to date

testing		been improved?		
Three months before end of project	How the work can be disseminated and best taken forward	Has the project achieved its aims? How can the outputs be used beyond the project? What other work has been suggested by the project?	Internal project staff evaluation	An exit strategy for the project (see below)

16.1 The success of the prototype and the demonstrator will be evaluated by the community user group.

16.2 The project anticipates a need to build upon initial findings to help improve the quality of PIPs, and also to take account of the evolving nature of PIPs through continued use of ticTOCs. As further ticTOCs usage data becomes available, this will influence and may re-shape PIPs, and the second round of testing will evaluate such developments. Members of the user group will be offered a retainer fee (£100 Amazon coupons – the costs of which will be funded by Cambridge Information Group), to be delivered after the second round of testing. The project will involve this group in a 'paper pilot' at the initial contact stage, which will outline the aims of the project and the proposed solution, and invite feedback.

16.3 A blog (who's RSS feed will be included in an appropriate database of RSS content) will be developed to keep the user group informed about progress, however it will be important not to expect too much informal involvement and also to avoid project 'fatigue' within the user group. In addition, ongoing and more frequent feedback will be invited from the U&I community in an iterative process. Whilst the general aims of the project are not expected to change, input from the U&I community may influence the project's technical specifications and development.

17. Quality Plan

Output	Prototype/demonstrator				
Timing	Quality criteria	QA method(s)	Evidence of compliance	Quality responsibilities	Quality tools (if applicable)
Every month	Is development meeting requirements?	Internal assessment	Email agreement by partners	All	
Ongoing	Is the software robust?	Internal software QA assessment	Report from technical manager to partners	SC	
End of project	Is the software doing what we said it would?	Peer review	Feedback from community	Hull	

Output	Testing				
Timing	Quality criteria	QA method(s)	Evidence of compliance	Quality responsibilities	Quality tools (if applicable)
After Phase 1 testing	Does Gold Dust tool meet user requirements?	User feedback	Feedback responses	Hull	
After Phase 2 testing	Does Gold Dust tool meet user requirements?	User feedback	Feedback responses	Hull	

Output		Specification for PIPs			
Timing	Quality criteria	QA method(s)	Evidence of compliance	Quality responsibilities	Quality tools (if applicable)
March 2009	Is the specification suitable for scaling up to live service?	Review against testing of the specification to date	User feedback on the quality of the PIPs	Hull/Heriot-Watt	

Output		Databases of aggregated feeds			
Timing	Quality criteria	QA method(s)	Evidence of compliance	Quality responsibilities	Quality tools (if applicable)
September 2008	Does the database contain the required feeds?	Review of available feeds	Updating of database	Heriot-Watt	
Dec 2008	What gaps exist in the database of feeds for effective delivery of Gold Dust?	User feedback on the quality of Gold Dust output, review of available RSS feeds	User feedback	Hull	

Output		Report on potential commercial applications			
Timing	Quality criteria	QA method(s)	Evidence of compliance	Quality responsibilities	Quality tools (if applicable)
March 2009	FFP	Review by commercial partners	Acceptance by commercial partners	Hull	

17.1 Gold Dust partners have already been involved in considerable user engagement activity of the kind identified by UIDM Stage 1. This has enabled the identification of issues and the development of an understanding of various specific information requirements and needs of the target user community.

17.2 An EEVL Engineering Portal Consultancy Group report, based on a user questionnaire, a follow-up survey and consultancy groups³³ identified many of the types of information included in the proposed Gold Dust databases of RSS content as being of particular interest to the community, and not covered adequately by existing information retrieval tools. Conference announcements, teaching and learning materials, articles in e-journals, news items, theses and dissertations and funding opportunities were all identified in this respect. The report also notes a desire for improved personalisation and alert (current awareness) capability. More recently, the JISC-funded PerX project conducted a web-based user questionnaire and focus groups which found³⁴ a need for easier access to current information, including funding information and improved alerting capabilities. This has been confirmed from other feedback received via the PerX project. Surveys and feedback conducted by Intute SET have also shown that users require 'rich' information, i.e. resources that provide information that is directly relevant to the query, that

³³ Additional Services for SPP: Summary of Portal Features Survey and Portal Consultancy Groups
<http://www.eevl.ac.uk/public/ASP/info/>

³⁴ PerX User Feedback report <http://www.icbl.hw.ac.uk/perx/userfeedback.htm>

is specific and informative, and that can be re-used, applied, or adapted according to the needs of those users. Feedback from a major Intute survey in October 2006 is available³⁵.

17.3 These stakeholder and user involvement findings are backed up by several of the reports cited in the background section of this document, which have themselves often been based on user consultations. Together, they confirm requirements of the broader community and a definite demand amongst users for better personalisation tools of the type envisaged by Gold Dust.

18. Dissemination Plan

18.1 The project partners have extensive experience of dissemination and will ensure that the project findings are adequately disseminated. Some Gold Dust concepts have been hinted at in articles already published^{36 37}. An invitation has been received from a popular publication for an article about Gold Dust, and an invitation has also been received to present at a European conference.

18.2 Additional publications will be targeted for appropriate dissemination, including an analysis of evaluation results, and we would hope to work with JISC Communications and Marketing with respect to further dissemination opportunities pertaining to the demonstrator. All documentation would be maintained beyond the life of the project. The full text of all published research papers and conference proceedings arising from Gold Dust work would be deposited in an appropriate open access repository.

Timing	Dissemination Activity	Audience	Purpose	Key Message
Project start	Set up project website and blog	Project partners Librarians JISC community Academics	To keep interested parties up-to-date with project activity and promote discussion around the issues being investigated	This is what the Gold Dust project is doing and thinking
Throughout the project	Gold Dust will link into ticTOCs dissemination activities as a specific follow-on from that project	Librarians	To demonstrate how the work of ticTOCs is informing Gold Dust	Using TOCs for added value personalised services
Month 9 to end of project	Conference presentations/demonstrations and an article (possibly joint with ticTOCs)	Librarians Academics	To present and/or demonstrate the outputs from Gold Dust and its work	Delivering personalised information to update you where it's needed
Phase 2 research and development	Surfacing of Gold Dust through varied presentation interfaces	End users Librarians	To demonstrate the capability of Gold Dust in a variety of settings	Adding value to existing interfaces through targeted information delivery
Post project	Presentations and demonstrations of implementations	Librarians Academics	As above	As above

³⁵ <http://www.intute.ac.uk/publications.html>

³⁶ R.A.MacLeod, RSS Update: It's RSS, Jim, but Not as We Know It, Free Pint, No 234, 26 July 2007, <http://www.freepint.com/issues/260707.htm#tips>

³⁷ R.A.MacLeod, Guru interview, Emerald LibraryLink, The Information Professional's Knowledge Network, April 2007, http://www.emeraldinsight.com/info/librarians/info/guru_macleod.jsp

19. Exit and Sustainability Plans

Project Outputs	Action for Take-up & Embedding	Action for Exit
Gold Dust prototype	Use by community user group	Positive evaluation leading to development of demonstrator
Gold Dust demonstrator	(a) Use by community user group	Positive evaluation leads to further development and implementation with a broader group of users, content providers, service providers and other institutional portals
	(b) sustainability of demonstrator	Extending the demonstrator to further institutional portals; in conjunction with OSS Watch
Partner knowledge	Adoption of the project's outputs in local environments and demonstration of use	

The project partners will ensure that deliverables are available beyond the life of the project.

Project Outputs	Why Sustainable	Scenarios for Taking Forward	Issues to Address
Gold Dust Demonstrator	If proven to be useful by a community user group then will be able to explore further development,	Avenues for further development and implementation, to the benefit of a broader group of users, content providers, service providers and other institutional portals, and including commercial partnership possibilities, can be explored	Scaling up the demonstrator to a full service

Appendices

Appendix A. Project Budget

NB: the project budget is not publicly available for reasons of privacy.