Citation for published version:
Ball, A 2010, 'The KiM and ERIM Projects' University of Bath Library Summer Staff Development Programme, Bath, 5/08/10, .

Publication date:
2010

Link to publication

University of Bath

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.
The KIM and ERIM Projects

Alex Ball

5th August 2010

My name’s Alex Ball and I work in UKOLN, for the most part with the Digital Curation Centre and associated projects. This will be a very brief tour through a couple of the projects I’ve worked on in collaboration with the IdMRC, namely KIM and ERIM.

1 KIM Project

KIM stands for Knowledge and Information Management through life.

- £5.5 million Grand Challenge project that ran over 3.5 years.
- Funded by EPSRC and ESRC.
- 80 industrial collaborators from aerospace, defence and construction.
- 13 partners across 11 universities.
- Knowledge, information and data management issues in the engineering industry.

Actually, the original title was ‘Immortal Information and Through Life Knowledge Management: Strategies and Tools for the Emerging Product-Service Paradigm’, but that didn’t fit very easily on a slide title.

1.1 The Product-Service Paradigm

But what is the product-service paradigm? It’s probably easiest if I illustrate it with an example.

Let’s say you have a publisher making electronic journals and putting them on CDs (Figure 1). The publisher sells the CDs to the library, who keep it in perpetuity. It’s the library’s job to make sure the CDs continue to perform the function they’re supposed to: stopping them getting damaged, providing a platform they run on, and so on. That is an example of the product-as-commodity paradigm.

Now consider a situation (reveal) where a publisher keeps hold of all the content, and instead gives the library access to that content for a fixed amount of time. This time it’s the publisher’s job to make sure that the content remains accessible and usable in perpetuity, and the library just has to keep the contract going. That is the product-service way of doing things.

Now, instead of electronic journals, imagine we’re talking about turbines, or planes, or tanks, or even buildings. Instead of buying the commodity, you pay someone to make sure you always have access to it. This is the kind of situation we were concerned with in KIM.
Figure 1: The product-service paradigm. Top: product as commodity. Bottom: product as service.

1.2 Project structure

The project itself was organized into four work packages.

   1.1 Extended product models.
   1.2 Information organisation, retrieval and visualisation.
   1.3 Automated information capture.

2. Learning Throughout the Product-Service Cycle.
   2.1 Information capture and feedback.
   2.2 Learning from use.
   2.3 Value of information.

   3.1 Commercial incentivization: espoused intentions and experienced realities.
   3.2 Human resource management.
   3.3 Life cycle decision support models.

4. Integrating Activities.
   4.1 Intellectual framework.
   4.2 Emerging outcomes.

The work package structure seemed quite a logical way of dividing up the research, but was a bit too rigid in its boundaries, and didn’t give a picture of how the research fitted together as a whole. We needed a better way of framing and integrating our activities. On advice from our industrial collaborators, we came up with the EGIPT Framework (Figure 2).

‘Practices’ was originally called ‘Techniques’, but as that would have made it the ‘Eejit’ Framework, we felt we had to change it.
Unfortunately I don’t have time to cover the whole project for you, so I’m going to concentrate on the things I’m most familiar with and which I think you’ll find most interesting. These will be mostly drawn from the Tools dimension, but some of them spill over into other dimensions.

### 1.3 Knowledge-Enhanced Notes

The first piece of research I want to pick out is an example of this. Most of the important design decisions in industry happen in design meetings. The University of Strathclyde built a system for recording these design meetings using a wiki (Figure 3). Digital documents, photos of sketches, etc. can be dragged and dropped onto the wiki page as they are used, and these can be synchronized later with audiovisual recordings of the meeting.

The reason for doing this is that it gives a much clearer picture of how decisions were made at the meeting. Mere written minutes – being rather coarse summaries – frequently fail to record decisions that may prove to be key to understanding the finished design. On the down side, the multimedia version is a lot more data to store, and takes a lot longer to analyse and sift through, so as a tool it would probably need to be used sparingly.

### 1.4 DRed

The next tool I want to show you is DRed, the Design Rationale Editor, which can be used to record the progress of both groups and individuals. It is a graphical rationale mapping tool based on gIBIS and developed by the University of Cambridge; it is also useful for root cause analysis. Questions are shown as question marks, possible answers are shown as lightbulbs, and pros and cons of the answer are shown as green pluses and red minuses.

The tool is also capable of bidirectional links between DRed maps, and between DRed maps and word processor documents, spreadsheets, images, etc. DRed can also import spreadsheet values directly into the map, to show how calculations affect an issue.
Belt or chain driven pulley system
Neither are geometrically constraining (EXP) (05.58)
Chain doesn’t deal with debris well (EXP) (06.07)
Belt has no lubrication maintenance issues (EXP) (06.19)

Figure 3: Knowledge-Enhanced Notes (KEN).

Figure 4: Example rationale map produced by Design Rationale Editor (DRed).
1.5 Automated information capture

To this end, Heriot-Watt University put together a system for analysing log files of design interactions to infer rationale and detect inefficiencies in the design environment (Figure 5). The research focussed initially on virtual reality design environments, such as that used for routing cables through an existing design. In order to apply the same principles to a CAD environment, the team used a modified version of the Bamzooki software produced by the BBC: http://www.bbc.co.uk/cbbc/bamzooki/. In this modified version, the designer is presented with a ready-made Zook (similar to the one pictured above) and given the task of improving it so that it can pass both a speed trial and a strength trial. This involves a number of design–test–revise cycles mirroring those found (albeit in a longer time frame) in real world design.

Motion studies techniques are used to analyse the log files produced by the design software. Of particular use are the chronocyclegraphs (tracking the motion of a worker’s hands) and therbligs (see http://gilbrethnetwork.tripod.com/herbligs.html) invented by Frank and Lillian Gilbreth. One of the easier things to determine is how much time a designer spends doing a particular type of activity; in one trial using the VR system, it was revealed that designers spent around 47% of their time navigating menus, indicating that the design interface needed to be improved. (Comparison between the Bamzooki logs and earlier studies of real-life CAD design also showed close correspondence, showing the validity of using Bamzooki as a testbed.) Analysis of the patterns of work is a little harder, but can be used to determine the more effective ways to go about solving a problem, and to infer the rationale behind certain design choices.
1.6 Topic maps

Helping to bring all this together was some work done by the University of Bath on topic maps. Topic maps link topics together across documents, and reveal information dependencies (Figure 6). They can be used to link together group and individual activity so that the entire design process is mapped out.

- **Topic**: piece of information, activity
- **Association**: input, output, control, resource
- **Occurrence**: underlying information resource

Individual work produces documents and ideas that are presented in design review meetings. The design review meetings set the agenda for future individual work. The topic maps explicitly record these information and process flows.

1.7 Faceted classification

The University of Bath also worked on a browsing interface for records based on facets (e.g. pumps classified by pumped material, mechanism, pressure).

These facets aren’t true facets, as they’re not strictly orthogonal, but it was felt that trying to be rigid about this would throw away some useful access points. The interface shows each facet beginning at the same hierarchical level (Figure 7). As selections are made within one facet, items that do not correspond with the selection are removed from the branches of the other facets. This way a designer can home in on the required item by making selections in the order that makes sense at the time.

This system was used for classifying reports on in-service maintenance activity. Fixed fields within the reports (part number, engineer’s name, etc.) are easy to represent as facets, but free text fields are more difficult. Text mining approaches were tried for finding useful facets from a corpus of reports, but results varied depending on how consistently the service engineers filled out the reports in terms of the language they used. If nothing else, this research showed that using more formal language in reports made it easier to discover systematic behaviours and trends in this way.
1.8 Information management using RFID

One of the strands of research at the University of Reading was proving the concept of using a shared database of components alongside Radio Frequency Identification (RFID) tags to enable through-life knowledge management.

The idea (Figure 8) is that there is a shared central database that holds all the parts available to a designer, and records of all the parts currently in use. The designer uses the database for selecting components (much as catalogues of parts are used today), and the finished design refers back to these parts in the database. When the design is realised, RFID tags are used to match up the real world components to the components in the design and the database, and are used to track the components throughout their life from supply to disposal. At any point in the lifecycle, RFID readers can be used to look up a component in the database, so that the corresponding record can be read and added to. Certain RFID tags can also act as sensors, providing real-time data on the performance of the finished design. Not only can this help to keep track of important information about components, such as whether they need to be disposed of in a particular way, it can also help with spotting systematic problems with certain types of component and comparing the performance of different designs.

1.9 Annotation of lightweight representations

Lastly, I want to talk about the areas I worked on within the KIM Project. Along with Lian Ding of the University of Bath IdMRC, I looked at making design documentation more curation-friendly.

Without going into too much detail about this, CAD models are a real problem for anything other than an immediate design task. The files themselves are more like source code that has to be run through an interpreter, and those interpreters are locked down, proprietary and mutually incompatible even between different versions of the same software. Furthermore, the file sizes are enormous and the files themselves are jealously
guarded intellectual property. Re-using that CAD data in future upgrades and redesign work, or even for manufacture and maintenance, is a very tricky operation.

Fortunately (reveal) there are rather more amenable formats out there that are simpler, with public specifications and interpreters. They don’t contain the full richness of the original model, but they contain enough to be useful for specific use cases. The proposal we had was to supplement these formats with separate annotation layers (reveal) that can be layered over the top. These can contain some of the missing CAD information, or might be used to feed information back to the design team on, say, how the design performs in practice. This information can be layered back over the original design as well (reveal), or used to help reconstruct a full CAD model.

We put together a proof-of-concept implementation of this called LiMMA (Lightweight Models with Multilayered Annotations), using the NX as the full CAD system and X3D and 3D PDF as lightweight surrogates.

Figure 8: Information management using RFID. Clockwise from left: parts database; supply; installation; performance in service; disposal. Bottom: Building Information Model (BIM).

Figure 9: Lightweight Models with Multilayered Annotations (LiMMA).
1.10 Representation Information

The choice of which lightweight format to use depends on what information you need it to keep and what you want it to lose, or don’t mind it losing. As an aid to making this choice (Figure 10), we put together a tool for determining the ideal surrogate format for the geometry, and the most suitable tools for generating it, determined by comparing user requirements for things like exact 3D surfaces or file streaming, and the support given by the various formats and converters.

2 ERIM Project

All this talk of managing and preserving data brings me round to the project that myself and Mansur are currently working on, and that is the ERIM Project.

- Engineering Research Information Management.
- Funded by JISC.
- Research Data Management Programme, Research Data Management Planning for Research Funders’ Projects strand, the funder in question being the EPSRC.
- University of Bath: IdMRC and UKOLN/DCC.
- Managing data produced by the KIM Project and other IdMRC research.
We’re about halfway through the project, and aside from the standard review of the state of the art, I suppose our main work of interest has been in two areas.

The first is a terminology for describing data, the relationships between them and the ways in which are manipulated (Figure 11). Some of the key ones are Data Purposing, which is preparing data for its intended purpose, Data Re-purposing, which is preparing data for another known use, and Supporting Data Re-use, which is preparing data for an unknown future use.

You can see some other ones in the word cloud like Collect, Aggregate, Augment, Derive and so on. These bring me to our second area: Research Activity Information Flows, or RAIFs (Figure 12). These RAIF maps are written in UML, and describe the processing steps from first-generation data to final data using the verbs I just mentioned. We’re now considering how researchers might be able to create diagrams like these as they go along, and whether any of this can be automated. The other major deliverable of the project will be example data management plans for engineering research data and a method for writing more.
Figure 12: A Research Activity Information Flow map (RAIF).