ABSTRACT

Several companies with experience in web development were interviewed about aspects of software engineering (SE) when developing Web applications. A semi-structured questionnaire was used to investigate their development processes, quality assurance (including metrics used) and tool support. The interviews are a first stage of empirical assessment. Among research observations identified are the lack of methods, activities and artefacts to reflect hypermedia characteristics (structure, content and navigation) of web applications. Another finding highlights the difficulties and lack of rigour when planning web-based applications because of the novelty of this type of application. In general, quality is assessed in an ad-hoc manner, without following any process or guidelines. Equally, reuse is also carried out in an ad-hoc manner and sometimes components are developed from scratch because developers do not trust the quality of the components or the sources. Last but not least, there is a lack of quantitative feedback to quality improvements of both processes and products related to web development. Future work will investigate some of the issues raised from the interviews. Investigations will consist of evaluating different hypermedia methods and processes to model Web applications (e.g., extending UML) by means of sound quality models such as ISO9216, ISO15504, etc. Further, our approach will incorporate quality metrics in the development processes by which could be automated and analysed to estimate and enhance quality characteristics of the software process itself and those of the product.

Keywords: Empirical Software Engineering, Web development, survey
1 Introduction
This document reports on a series of semi-structured interviews, which have been conducted with web-based applications developers/managers aimed at analysing the processes involved in the development of Web based applications.

The aim of the survey was to explore the experiences of practitioners directly in order to discover the main issues involved in the development of web applications. In particular, issues surrounding the activities, artefacts and metrics needed when developing these kinds of applications. It was hoped that significant points would be raised that would warrant further empirical assessment.

This paper is organized as follows. In section 2, the types of empirical assessment are described and the contribution that each of them makes to the establishment or investigation of a hypothesis. Section 3 explains the design, context and conduct of this survey; In section 4 the findings of the structured interviews are described. In section 5, conclusions and finally in section 6, future work based on the results of the interviews are presented.

2 Empirical Assessment
So far, very little quantitative experimentation has been performed in areas like web and hypermedia development. In order to develop better web applications, quantitative experimentation needs to be performed for validating the current methods, models, concepts, artefacts, etc. The general lack of empirical validation of software engineering (SE) theories is described in [6, 7] and [9], and the authors are of the view that SE research has become insular and academic. The vision of Glass is summarised in three points:

- Software practice and research work should together
- Good research results should be integrated into practice via an established process
- Bad research ideas should be discarded quickly and not kept alive purely by research advocates

In addition, as Perry et al. [16] suggest, empirical studies can be used not only retrospectively to validate ideas, but also proactively to direct our research, identifying and justifying the potential value of research.

2.1 Empirical Assessment Frameworks
Fenton and Pfleeger [7] consider three ways of organising an evaluation exercise: survey, case study or formal experiment:

- **Survey** is a retrospective study of a situation to try to document relationships and outcomes.
- **Case study** is a research technique where the investigator identifies key factors that may affect the outcome of an activity and then documents the activity.
- **Formal experiment** is a rigorous, controlled investigation of an activity, where key factors are identified and manipulated to document their effects on the outcome.

Formal experiments, case studies and surveys can be done either quantitatively or qualitatively. Kitchenham [11] describe a method for selecting evaluation techniques for SE as part of the DESMET project. Daly [5] points out the value of using all three forms of empirical assessment to support each other in generating a hypothesis and establishing the results. The qualitative survey contributes to the formulation of a relevant hypothesis, formal experiments confirm if a relationship exists with the hypothesis, and finally, case studies determine whether the results can be generalised.

2.2 Qualitative Evaluations via Interviews
According to Seaman [19], interviews are a qualitative method in empirical studies used to collect historical data from the memories of the interviewees. They can be of several types: structured, unstructured or semi-structured. A structured interview consists of an interview where "the questions are in the hands of the interviewer". In unstructured interviews, "the interviewee is the source of both questions and answers" and the objective is to get as much information as possible from a topic. Semi-structure interviews consist of open-ended questions to elicit both the information foreseen and unexpected questions.

The analysis of the qualitative data can be twofold: generation of theory or confirmation of a theory [19]. Generate a theory is performed extracting from a set of field notes a statement or preposition that is supported by the data. Basili et al. [1] also state that surveys contribute to the formulation of the hypothesis and increase the likelihood of studying what is relevant.

In addition, empirical assessment techniques need to be used to check the results of the interviews. In order to generate a theory, the constant comparison method [8, 19] consist of adding codes to field memos, then group information according to the codes and writing a field memo synthesising the findings.
3 Design of the Survey
We performed a survey in industry about the development of web applications. The purpose was to get feedback, from experienced individuals, on issues relating to the quality of the resulting products as well as the processes.

3.1 Aim
The aim of the survey was to explore the experiences of practitioners directly in order to discover the main issues involved in the development of web applications. In particular, we wanted to identify issues surrounding the activities, artefacts and metrics that are needed when developing these kinds of applications. The survey is seen as the first stage of a multi-method approach to research in empirical SE as described by Daly [5]. The purpose of this first stage is to raise interesting and relevant findings to contribute to the formulation of the hypothesis and identify what is relevant for further research [1].

3.2 Design of the Questionnaire
In order to cover a broad range of topics the survey was conducted by semi-structured interviews based on a questionnaire presented in appendix A. The questionnaire was structured following the software process life cycle in order to facilitate the analysis and coding of the results following empirical assessment techniques. Members of the SE group at the university reviewed the questionnaire and their feedback prompted some changes to the questionnaire concerning ambiguous questions and how to focus of the questionnaire.

3.3 Selection of Subjects
There is always a trade-off between an extensive but superficial survey and an in-depth but possibly unrepresentative survey. With a small set, a wider exploration of the topic and a higher level of confidence in the answers is possible. We tried to find successful companies involved in web applications in order to get subjects with experience in web development. Companies of different size and type were selected.

3.4 Conduct of the Interviews
The companies were initially contacted by email with a brief outline of our aims and the questions that would be asked. The interviews were structured around an open-ended questionnaire but the interviewees were encouraged to digress and elaborate on topics as much as they felt necessary.

<table>
<thead>
<tr>
<th>ID</th>
<th>Type of Web Apps</th>
<th>Type of interviewee</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Web IS (3 interviews/2 dep’t.)</td>
<td>Software engineers</td>
</tr>
<tr>
<td>B</td>
<td>Software Systems</td>
<td>Project manager</td>
</tr>
<tr>
<td>C</td>
<td></td>
<td>Sales consultant</td>
</tr>
<tr>
<td>D</td>
<td>ISP and Web design</td>
<td>Software engineer</td>
</tr>
<tr>
<td>E</td>
<td>Insurance Web solutions</td>
<td>Software engineer</td>
</tr>
<tr>
<td>F</td>
<td>Financial and Pharmaceutical Web solutions</td>
<td>Project manager</td>
</tr>
<tr>
<td>G</td>
<td>Web Design</td>
<td>Web designer</td>
</tr>
<tr>
<td>H</td>
<td>Government IS</td>
<td>Project manager</td>
</tr>
</tbody>
</table>

Table 1: Companies interviewed

4 Results of the Survey
It must be borne in mind that the results were obtained from a limited sample. The results should be viewed as a starting point for further empirical research.

4.1 Project Management
The general view was that Web projects are very similar to all other software projects, except that toolsets for development are used. The pace of change, the reduced time to market, fast turnaround and prototyping add challenges to management. For example, an interviewee in the organization A commented that "a problem that has frequently occurred is that a prototype has been rushed into production either at the customer’s insistence, or at the wish of our own management, with lots of resultant problems". Our view of these problems is that they are not specific to web development but they are more severe than other application domains. Some of interviewees noted that more research in project management would be beneficial, but came from a manager not involved only in technical roles.

4.1.1 Estimation
Surveyed companies estimate schedules, cost and resources by analogy using judgement and experience from previous projects, adding some risk factor depending on the use of novel technologies (e.g., using XML for first time). The interviewees commented that such estimation may be very difficult. The estimates are more accurate after a few similar projects have been carried out. The common technique is to create a WBS (Work Breakdown Structure) as accurately as possible, which in turn helps in plan estimation, milestones setting, and resources assignment. From our point of view, this lack of process estimation baselines means that there is a lack of rigor in some process activities.

4.1.2 Process Metrics
Most of organizations surveyed collect metrics related to time and effort by means of timesheets, which are used for billing the clients and/or for better estimations in future projects. Other common measures in large organizations are related to maintenance, such as the phone calls received by help desks and the bug fixes when the system is in service, as this contributes to the users' perception of the usefulness of the system. A well-defined pattern observed from the surveyed companies is that processes are better measured in larger companies.

4.1.3 Resources
A pattern that arises from the interviews is that, in general, there is a clear distinction of front- and back-end development. According to an interviewee in organization A, this is very beneficial in separating presentation from functionality. In organization H, which is a large organization, developers in the front-end only have the high-level view of the application. They need to know about the high-level functions to call, for example, ASP pages that call functions in the backend processing. On the other hand, back-end developers work at the object level. The interviewee commented that this was a quite different from traditional systems where developers are more homogeneous. In the smaller companies where applications are static sites or small database driven applications, the distinction is between the authoring/design role and the technical role.

An interviewee in organization B commented that more technologies are involved in web-based applications (Java, JavaScript, HTML, CORBA, etc.) than the traditional ones. Thus, more roles are involved and staff with more skills is required.

4.2 Requirements
The most common way of eliciting requirements is by means of use cases, which are used by the companies B, C and F. Company C uses standard questionnaires based on their methodology as well. Requirements are generally recorded just with documents and no special tools are used. However, company B uses Rational Suite to support requirements needs. In addition, prototypes and developmental releases of web sites are used in companies D, F and sometimes in A to match the client’s required functionality.

4.3 Analysis and Design
There are new activities and roles to manage the content, design and structure of web-based applications. However, approaches found in the academic literature for the design and development of Web applications were not used in practice.

4.3.1 Hypermedia Design in Web applications
Some researchers have proposed methods such as the Object-Oriented Hypermedia Design Methodology (OOHDM) [18] and Relationship Management Methodology (RMM) [10] in response to the design demands. The benefit of these methodologies is that they permit new activities and artefacts to be incorporated in the life cycle of SE.

It was observed from the data collected in the interviews that, no methodologies related to hypermedia, such as OOHDM or RMM, were followed. The most common technique was to create flowcharts and storyboards of navigation using a top-down approach. Company C used patterns as a starting point. In comparison, company B and H defined prescriptive tasks, which constitutes procedures and guidelines to be followed. Rational Unified Process with UML and activity diagrams were used in company B to model the hypermedia structure. Only one interviewee commented that new artefacts would be beneficial in representing the hypermedia domain. He commented that current methods work well when the number of relationships is small, but they are poor notations for modelling large web sites. In the literature, it is also commented by [15].

It is clear that new activities and artefacts are needed and should be incorporated into hypermedia engineering. Established SE techniques are still needed when developing web-based application, but processes should take into account the differences in Web applications. Researchers have recognized that Web-based applications are a distinctive class of information system [10, 13]. By extending UML as a notation language and by creating new activities to reflect hypermedia, it would be beneficial for creating more usable and maintainable Web applications. There are works enriching UML for hypermedia design such as Conallen [4] and Baumeister et al. [2] and support exist, to some extent, in current tools (e.g., Rational Rose enrich UML by means of stereotypes.
to permit modelling of Web architectural). This is an active area of research within the Hypertext Functionality (HTF) community, where researchers are looking at ways to incorporate HTF into application designs [3].

4.4 Metrics and Quality Assurance

4.4.1 Standard Certification
Some departments in organization A, and organizations B, C, D have ISO certification. Organization B is expected to reach CMM Level 2 in September. Regarding standard certification, one of the interviewees suggested that ISO certification does not necessarily mean a good and well defined development process, and having defined key process areas, like in CMM can be more useful. From the data collected, it is clear that the bigger the company, the more standards and procedures are defined and for example, configuration management and reuse are more ad-hoc in small organizations than in the large ones.

In general, quality is only seen in two ways: the absence of errors and customer satisfaction. Quality is really acceptability to the user and the performance quality is that which the user is happy with.

4.4.2 Product Metrics
The use of product metrics is quite sporadic in industry and in general, a measurement program is seen with little respect. The two main reasons raised from the interviews are that metrics are not found to be good value for money. Further, some interviewees were of the opinion that metrics pertaining to product quality are simply common sense. One interviewee at organization A commented: "At various times I have worked in groups where they were tried, but they were not found to be value for money".

Two of the interviews commented that quality metrics judgements are really a matter of common sense and this seems to be especially true with web systems. One of them used the readability example where readability is seen as a very subjective measure. Another commented on the example of user interface performance, where downloads are at the mercy of the intervening networks. Company B measures the basic attributes such as use cases, lines of code, etc. When the systems are in production, the metrics collected by the web server are extensively used to analyse the users. In general, few metrics are collected from the product when developing and these metrics are the ones that are easy to collect and automated, i.e., broken links, page size, etc.

4.5 Reuse
Only company C (the largest one) is using a process for reuse and supported by the use of tools (e.g. extensive Lotus Notes databases of work products for previous projects). Company B is implementing a reuse process and nowadays and there is big emphasis in document artefacts to be reused in the future.

The interviewee pointed out that reuse within a process is very important however sometimes components are developed from scratch because they do not trust the quality of the components. In addition, company F commented that they only buy components from trusted sources, i.e., solid companies that are expected to be in the market a for long time. Another interviewee in organization A commented: "It is really important when choosing the servers and software that you know that the people who support them are reliable, experienced with the product and not likely to change job (internally or externally) for the next few years. This is why it is often better from the user's point of view to have older technology used rather than the latest thing".

Reuse within the smallest companies (D and G) is also high but ad-hoc. This is because many of the applications or the structures of web sites are very similar. In addition, many applications can be obtained from the web itself as the work is customizing and integrating them but this knowledge is dispersed around the people in the company without following any process. An interviewee from A commented that the constant adoption of new technologies makes it difficult to reuse any pre-existing material. Some of the design patterns ideas can be incorporated, but everything else needs to renewed.

There is a different treatment for different parts of the applications. Frameworks and components that will be reused are developed more carefully and more attention is paid to them.

4.6 Maintenance
The organizations interviewed use configuration management procedures and tools. Usually sites are database driven and toolsets are used to manage links (for example IBM Websphere Studio). Company C is using their own tools and the Tivoli suite of products for configuration management. Company B is using the Rational Suite TestStudio and TeamTest. Web applications based on applets are tested as any other type of application because there is no difference between the development of applets and the development of traditional applications.

As we have seen above, metrics related to the web server's logs are also used for maintenance, checking the use of the system, broken links, type of users, and so on.

For example in the organization A, occasionally they measure maximum turnaround times for user queries and bug fixes when the system is in service, as this is what will contribute to the user's perception of the usefulness of the system. Organization H also collects metrics from the help desk.
4.7 Technology

4.7.1 From the Developers’ Point of View
In one case (organization A) the project was encouraged to use a integrated package for the web development and presentation because of the underlying application database. These integrated packages provided limited functionality that sufficed as long as no extra ‘out-of-the box’ functionality is required. When this happens, “we are continually fighting against some severe limitations and restrictions”. The interviewee recommended avoiding using such integrated tools in future and opting for more general pluggable systems.

In general, device independence is considered highly important in order to get universal user interfaces and cross-platform independence but a minimum is assumed (e.g. versions 4 of Netscape and Internet Explorer). It is almost true that keeping within the "low common denominator" and standards, the same document can be displayed on any client, not requiring different versions of GUI. Another benefit of keeping familiar user interface is that users are used to it and know how to use it. In addition, two of the organizations (E and F) have projects that involve the use of WAP technology and other devices such as kiosks. In these cases, user interface tests are hard because the number of versions and systems grows up exponentially.

4.7.2 From the Users’ Point of View
An interviewee from organization A commented "always remember that the customers do not get excited about novel technology, just about doing the job they are there to do better, faster and more easily".

Regarding customer satisfaction, an interviewee in organization A commented: "The ultimate quality of a system is judged by whether the end user uses it regularly instead of their previous method of working, as it speeds up and facilitates their own job, rather than because they are forced to by lack of alternative or management pressure. It is easy to write a beautiful, theoretically perfect system that the users just do not find attractive enough to use. In this case, though technically successful, the project is a failure".

4.8 Social Aspects
The final question asked the interviewee if there were any important issues that had not been covered. In most cases, the interviewee recapped some of the more important issues at this point but two of them raised the following social issues:

- Training users is much easier because users are familiar with the UI of the web and only need to understand the underlying data or some additional functionality. Much of this can be done by means of on-line help.
- There are social changes in the organization: more democratisation of the information (more information is more accessible to more people). The drawback is that the user becomes more anonymous, because it is more difficult to know who and where the users are.

5 Conclusions
These surveys have pinpointed certain research vacuums, e.g. where industry has not adopted current techniques for hypermedia design, or quality measurement by means of metrics plans. The general feeling from the survey was that there is no significant difference between web-based systems and other systems from the management point of view and the same SE principles are followed.

Some interviewees agree that new activities (e.g. authoring), extensions to existing activities (e.g. link testing) and new artefacts are needed for developing Web-based applications. Although some hypermedia methods have been developed in the last decade, they are still far away from being used in industry. Authors suggest that possible reasons are:

- The hypermedia models/concepts are not clearly defined. Therefore, there is no common standard to represent these concepts and every hypermedia methodology (HDM, OOHDM, RMM) does it in a different way.
- Hypermedia methods such as OOHDM and RMM focus on the design and development of hypermedia applications but there are no guidelines to incorporate them into a general development process. Thus, there are no clear rules about how apply the hypermedia metaphor to the domain. This area of research is being carried out in HTF field [3].
- In addition, hypermedia methods can be too complicated for the average web project.

In general, quality is assessed in an ad-hoc manner, without following any process or guidelines. This means that there is no visibility from the process point of view about the quality of the developed product. This seems to be a common view in measurement programs where they are seen with little respect and with no return on the investment of time spent in data collection. No metrics are computed when they require heavy manual computation; thus, only metrics that are easy to collect or are automated by tools are used.

Equally, reuse is also carried out in an ad-hoc manner. This means that sometimes components are developed from scratch because developers do not trust the quality of the components or the sources. Without reuse
processes or guidelines in the acquisition, reuse is far from being close to the optimum levels. There are already techniques that are not extensively applied in the industry to mitigate this problem. Meyer [14] propose design-by-contract as a way to improve correctness and robustness of object-oriented modules by means of formal agreements between a class and its clients: “Only through such a precise definition of every module’s claims and responsibilities can we hope to attain a significant degree of trust in large software systems”.

6 Future Work
The results of the interviews are a useful input for further investigations. The following points are possible research issues.

6.1 Experiments into activities and models associated with modelling hypermedia
In order to perform experiments into modelling web applications, a critical review of the hypermedia process and metrics has been performed [12, 17]. Analysed hypermedia methods consider the following activities with different approaches:
- Define the application domain model. This step consists of the identification of interesting domain elements and their relationships (e.g. UML models, E-R models, etc.).
- Define the hypermedia domain model. The objective is to define hypermedia elements that are suited to the hypermedia metaphor (e.g. nodes, links, landmarks, guided tours, etc.)
- Definition of the relationships between application and hypermedia domains.

Further empirical investigations will consist of evaluating different hypermedia methods and approaches to model Web applications, e.g., extending UML to represent hypermedia functionality [2].

6.2 Ways to Incorporate Quality Measurement in Projects
Our approach is to incorporate quality metrics in the development processes such that they could be automated and analysed; i.e.
- Improve the development process of these types of applications from the process point of view (how to plan activities and its order, estimation, etc for improved management).
- Define a quality model for web development activities and products in a way that is easy to use and supported by means of tools.
- Automated tools for collecting metrics are therefore desirable. In addition, authors plan to apply techniques such as Bayesian Belief Networks (BBN), data mining etc. to help to estimate quality characteristics of both the software process and the product.

7 References
Acknowledgements
Thanks to all the companies and participants in these interviews who made this possible. To all members of the Applied Software Engineering group at the University of Reading for their suggestions and ideas and the University itself for its financial support.

Appendix A – Questionnaire

General Information
1) With respect to your company:
   a. What is the nature of your organization?
   b. How many employees does your organization have? And the usual web project development team?
   c. How much experience (in years) does your company have in web development for itself? For others?
2) Are the systems for internal or external users to your organization? (or in other words, what type of applications do you develop -intranets, extranets or for the Web-?)

Requirements
3) Which methods and metrics do you use to estimate the development effort of web applications (time, effort, cost)?
4) What tools are used when managing requirements?

Design and Development
5) Which development methods are used by your organisation when developing web applications? (e.g. Structured A/D, OO A/D methods, Data driven methods)
   a. Do you use any hypermedia specific methods such as OOHDM, RMM or HDM?
   b. How do you model the hierarchy and design of the web applications? (eg UI Patterns, storyboards, UI testing)? Please state the work products used in analysis and design to reflect hypermedia and navigation (eg storyboards, state charts, etc).
   c. Is device independence (browser, TV, phones, ...) considered?
6) What kinds of tools are used when developing these types of applications?

System test and integration
7) How do you test these types of applications?
8) What tools are used when testing?

Maintenance/Evolution
9) How do you maintain web applications with respect to the application executables (CGI, ASP, Servlets, etc.)?
10) How do you maintain the content and navigational structure (content and links)?
11) Does your organization have Configuration Management procedures? If so, please state which tools and techniques are used.

Reusability
12) Do you reuse application executables from previous applications? If so,
   a. What do you reuse (components/frameworks, documentation...)?
   b. How is the component acquisition process/development fitted into the software life cycle? How do you reuse it (ad-hoc, within a process, ...)?
13) Do you reuse data (content) from previous systems? If so,
   a. How do you reuse the content (granularity of the information, etc)?
14) What tools are used for reusing?

Metrics and Quality Assurance
15) Does your organization have a quality standard certification? (eg ISO, CMM, etc.) Quality Management team? Standard procedures and project standards?
16) Are metrics collected at your organization (system development metrics and process metrics)?
   a. If so, please list these metrics and why are they collected?
i. as part of the you life-cycle process
ii. as part of the web application product (eg size metrics)

17) Do you define quality objectives for each project? If so,
   a. How do you define these quality objectives?
   b. Are the ranges of acceptable values for each quality characteristic defined? (eg. size limits, page layout, etc.)?
   c. Do you consider different quality characteristics for different parts/artefacts of the application?

Project Management

18) How is the typical organization of your web project? Are there differences with other types of projects? If so, can you state them (activities, models, roles, …)

19) How do you manage the authoring process (content) and the development process (application)?

20) Is an iterative and incremental process followed?

Others

21) Do you have any comments not covered so far?