RESEARCH ARTICLE

The cathedral and the bazaar of e-repository development: encouraging community engagement with moving pictures and sound

Denis Wong, Kerry L. Shephard and Peter Phillips

Information Systems Services, University of Southampton, UK; Higher Education Development Centre, University of Otago, New Zealand

This paper offers an insight into the development, use and governance of e-repositories for learning and teaching, illustrated by Eric Raymond’s bazaar and cathedral analogies and by a comparison of collection strategies that focus on content coverage or on the needs of users. It addresses in particular the processes that encourage and achieve community engagement. This insight is illustrated by one particular e-repository, the Education Media On-Line (EMOL) service. This paper draws analogies between the bazaar approach for open source software development and its possibilities for developing e-repositories for learning and teaching. It suggests in particular that the development, use and evaluation of online moving pictures and sound objects for learning and teaching can benefit greatly from the community engagement lessons provided by the development, use and evaluation of open source software. Such lessons can be underpinned by experience in the area of learning resource collections, where repositories have been classified as ‘collections-based’ or ‘user-based’. Lessons from the open source movement may inform the development of e-repositories such as EMOL in the future.

Keywords: repositories; community; video; audio; collections; open source

Introduction

Raymond’s (2000) ‘The cathedral and the bazaar’ was a description of the open source software movement. Perhaps more significantly, it was a deep reflection upon a way of doing things, questioning conventional wisdom in process management and analysing processes that were previously taken for granted. Despite its informal style, it was an in-depth dissection of a world class open source product, the Linux operating system. The reflection was inevitably in hindsight, but with lessons both specific and general. So, for example, it highlighted specific management styles that led to highly cost effective, specific code production. At the same time it indicated how the application of such approaches could be generalized and used in other areas of endeavour, not just software coding. This paper offers a case study based insight into one such area, that of e-repositories. It provides a critique on the development and governance of e-repositories in general, but based in particular on one that the authors had experience of; the Education Media On-Line (EMOL) service.

The focus of this paper is not upon the delivery platform. Rather, it is the area of community engagement with e-repositories and of finding ways so that they achieve critical mass and become useful tools for their users. It suggests changes in management emphasis so that services like EMOL might achieve not only quality in technical service delivery, but genuine improvements in the usability and use of the content collected, with all of its potential for improved learning and teaching.

*Corresponding author. Email: d.wong@soton.ac.uk
Background: EMOL as an e-repository

The Education Media On-Line service (EMOL, now the Film and Sound On-Line service) of the UK’s Joint Information Systems Committee (JISC) delivered high quality moving pictures and sound (MPS) objects through an online portal (EMOL archives are available at http://www.filmandsound.ac.uk/maas/). Phillips, Shephard, and Wong (2005) carried out a review of this service on behalf of JISC. The service at that time consisted of 16 collections, containing hundreds of titles and many days worth of viewing material. For example, the Royal Mail Classics collection contained 63 titles that ran for 14 hours; the Sheffield University Television collection had 111 titles that ran for 45 hours. The aims of that review were essentially to comment on the current policy on development and governance of that particular electronic repository and provide future direction to this collections policy.

Although the collection’s precise policy on development and governance was not clearly described by either EMOL or by the Managing Agent (http://www.filmandsound.ac.uk/description/maas.html), Phillips, Shephard, and Wong (2005) showed that the existing service was the result of a collections-based process which Calverley and Shephard (2003) described as ‘building collections that focus on content coverage’. In others words, material is gathered according to established ‘expert opinion’. Some of this material might be acquired at great cost, but often it is according to what is readily obtainable. This can lead to a ‘mismatch between what people are looking for (or expect) and the availability of resources that meet given quality criteria’, and factors beyond their immediate control may make this inevitable. For example, there may be a lack of available material in a popular area, expectations of quality might change dramatically over short periods of time and complex rights issues (particularly affecting MPS resources) may obstruct procurement of the desired material. The corollary is that with a collections-based process, material may be little used. A prior study by Griffiths (2005) indicated that the EMOL service suffered from this. Despite undoubted quality within the collection, the benefit in relation to cost was low.

Phillips, Shephard, and Wong therefore underpinned their study with the premise that the EMOL resource collection should move towards a user-based process, or what Calverley and Shephard described as ‘building collections that focus on users’. They commented:

Needs-analysis is a central feature of this process, and can be placed firmly within the discipline of educational evaluation. Traditionally, activities within this discipline have involved educational developers and educational evaluators supporting the activities of teachers. Needs-analysis also has a place within what must be more properly described as ‘market research’ or even ‘marketing’, as it is often necessary to introduce new resources to academic staff before they will consider using them. Both are less traditional roles for education professionals.

The user-based process therefore involves active, not passive, support of the teacher. Calverley and Shephard (2003) argued that needs analysis goes far beyond a simple identification of a potentially useful resource. To feed useful information back into the process of building collections it is important to know how the resource was used and how successful it was at supporting learning. The processes of needs analysis, therefore, involve a broad exploration of both user needs and user satisfaction. These authors argued that this approach brings new challenges for all involved in the creation, provision and embedding of resources within the learning process.

This process becomes particularly important as we move away from library building and a print base as the major learning resource (Roush 2005; Gandel, Katz, and Metros 2004; Wainwright 2004) and toward e-libraries, e-collections and e-repositories. It is within this context, characterized not only by a change in medium but also by a move towards almost unlimited globalized content, that contrasting content-based and user-based collection processes should be evaluated. The analysis of EMOL presented in this paper allows us to comment on these
considerations in this particular case. In the rest of this paper we will consider how a particular model (or analogy) can be applied to both EMOL and other collections, to describe how users and resources interact, develop and are governed in an online environment.

The cathedral and bazaar

A comment made by Raymond in 2000 managed to capture the mood of the open source movement at the time:

Who would have thought even five years ago that a world-class operating system could coalesce as if by magic out of part-time hacking by several thousand developers scattered all over the planet, connected only by the tenuous strands of the Internet? (Raymond 2000)

This illustrates what Raymond meant by a ‘bazaar’ – an open approach through which, in this case, software code was made ‘open’ and available for anyone to see, with the implication that no direct financial gain was to be made either by the original author or by those who joined him/her. Remarkable achievements, it seems, are possible with this approach, which can be contrasted with those of the ‘cathedral’, epitomized by the closed, managed approach, for example, of the Microsoft Windows operating system, which may be perceived not only in terms of a dominant desktop system but also the management methods behind it. With newly evolving bazaar approaches within the software industry such an approach is no longer the only way to produce software, and this could soon be the case for other products and areas as well.

Raymond described his bazaar in terms of 19 lessons (see Figure 1), illustrated by examples from Linus Torvalds’ Linux development. His cornerstone is described as ‘Given enough eyeballs, all bugs are shallow’ or, more formally, ‘Given a large enough beta-tester and co-developer base, almost every problem will be characterized quickly and the fix obvious to someone’. He wrote:

In the cathedral-builder view of programming, bugs and development problems are tricky, insidious, deep phenomena. It takes months of scrutiny by a dedicated few to develop confidence that you’ve winkled them all out. Thus the long release intervals, and the inevitable disappointment when long-awaited releases are not perfect. In the bazaar view, on the other hand, you assume that bugs are generally shallow phenomena – or, at least, that they turn shallow pretty quickly when exposed to a thousand eager co-developers pounding on every single new release.

The cathedral model uses project management to deliver defined goals to maintain a common direction, monitoring to ensure crucial details are dealt with, motivation to ensure productivity and organization to sustain the project. To build the cathedral, which is initially only a mental construct, each has to be manufactured and the workers have to be convinced, by religious, financial, political, social or other means, to become a part of the project. Construction is hierarchical and top-down. In the bazaar approach, on the other hand, we take a pre-existing social network, participate in it, learn about it and then stimulate it into a coordinated, focused and directed activity. Personal interest is a primary motivating factor, as is a deep respect for peer pressure. The underpinning means (Internet mail, pre-existing friendships, blogs, etc.) to stimulate each of these on a large scale are clearly also key. With both of these in place, the pre-existing network and the means for stimulating it, it is then a process of management through a ‘take it or leave it’ approach. The common direction evolves as a process of Darwin-like natural selection, with those interested staying on board (the survivors) and those not interested simply departing (at least temporarily) from the scene.

Frederick Taylor’s ‘cathedral’ view was propounded early in the 20th century during a period of corporate rise and the expansion of large scale industry. In his seminal book on scientific management he attempted to address massive inefficiencies in almost all human activities and to
1. Every good work of software starts by scratching a developer's personal itch.
2. Good programmers know what to write. Great ones know what to rewrite (and reuse).
3. "Plan to throw one away; you will, anyhow". (Fred Brooks, The Mythical Man-Month, Chapter 11)
4. If you have the right attitude, interesting problems will find you.
5. When you lose interest in a program, your last duty to it is to hand it off to a competent successor.
6. Treating your users as co-developers is your least-hassle route to rapid code improvement and effective debugging.
8. Given a large enough beta-tester and co-developer base, almost every problem will be characterized quickly and the fix obvious to someone. Or, less formally, "Given enough eyeballs, all bugs are shallow." (Linus's Law)
9. Smart data structures and dumb code works a lot better than the other way around.
10. If you treat your beta-testers as if they're your most valuable resource, they will respond by becoming your most valuable resource.
11. The next best thing to having good ideas is recognizing good ideas from your users. Sometimes the latter is better.
12. Often, the most striking and innovative solutions come from realizing that your concept of the problem was wrong.
13. "Perfection (in design) is achieved not when there is nothing more to add, but rather when there is nothing more to take away."
14. Any tool should be useful in the expected way, but a truly great tool lends itself to uses you never expected.
15. When writing gateway software of any kind, take pains to disturb the data stream as little as possible and never throw away information unless the recipient forces you to!
16. When your language is nowhere near Turing-complete, syntactic sugar can be your friend.
17. A security system is only as secure as its secret. Beware of pseudo-secrets.
18. To solve an interesting problem, start by finding a problem that is interesting to you.
19. Provided the development coordinator has a communications medium at least as good as the Internet, and knows how to lead without coercion, many heads are inevitably better than one.

Figure 1. Eric Raymond's lessons of the bazaar (summarized from Raymond 2000).

convince the reader that the remedy for this inefficiency lay in systematic and scientific management of these activities (Taylor 1911). Raymond's lessons of 2000 could hardly be more different – not so much in terms of particular work processes but, instead, in how those processes are controlled. Significantly, Raymond's lessons are tied to particular attitudes, motivations and
mindsets, constituting underpinning principles that might be independent of the type of work being done. Differences are in the conception of work, both in its fine grained control and its end product, which, in the case of open methods, might not be one packaged item but an evolving series of them. For example, the lesson that ‘Every good work of software starts by scratching a developer’s personal itch’ is about a hacker’s (open source programmer’s) sense of personal satisfaction; the cornerstone lesson that ‘Given enough eyeballs, all bugs are shallow’ is about voluntary actions and peer effects within a healthy, considerate community. Control, therefore, sits not with an overarching management hierarchy but within the programmer’s intrinsic motivation and with his/her sense of community. Jeremy Bentham’s 18th century panopticon (circular prison with cells distributed around a central surveillance station) is turned inside out, with those on the inside not just doing the work but supervising (or self-supervising) as well (Bentham 1787).

Through these underpinning principles extraordinary achievements arise. It is possible, therefore, to describe ‘Torvalds’ (and others’) pragmatic software development as a macro-phenomenon in which simple ideas evolve into complex systems (not just software, but matching social systems as well) that constantly renew, self-sustain, manage resources and fit neatly into ecological niches within a wider surrounding environment (financial, political, etc.). Kuwabara (2000), for example, compared Linux development with the biological evolution of the eye, for which Dawkins and others vehemently argue that no ‘intelligent design’ was necessary (Dawkins 1990; Coyne 2005). He claimed that the eye was an emergent phenomenon insofar as its eventual form was ‘statistically improbable in a direction that is specified without hindsight’. Similarly, when Linus Torvalds decided to design a new operating system (‘nothing big and professional’) he simply wanted a Unix system for his PC. He had little idea that it would develop into a global development project and a potential challenger to Microsoft Windows (http://www.linux.org/info/linus.html).

Using this perspective, Mulgan, Salem, and Steinberg (2005) argued that open source software development is only one instance of a broader methodology, for which they coined the term ‘open methods’. In particular, they saw these methods as applying to Wikipedia, an example of open knowledge. Founded in 2001 by Jimmy Wales (Pink 2005), it is an online encyclopaedia which has been built up from ‘stubs’, short pages of text to which users can add text of their own for moderation and update. To date it has almost 2,000,000 articles in English (August 2007) and many more in over 100 different languages. Such an idea could hardly be contained and ‘wikis’ are now standard types of ‘social software’ (Shirky 2003) in which all kinds of content, not just encyclopaedias, can be evolved using an open knowledge approach (Fountain 2005). With open source software the end product can be high quality, good value software. With open knowledge the end product can be high quality, highly usable content (Giles 2005). In both cases the process matters every bit as much as, if not more than, the end product. Associated with the open method or bazaar approach is a sense of community whose characteristics include:

- intrinsic motivation amongst workers for whom participation matters;
- reusability not just of product, but skills and knowledge as well;
- practicality and usability of the end product, because the communication gap between developer and end user is bridged.

The e-repository as cathedral

Cathedrals are large and imposing – they represent the ideologies and power structures extant at the time of building. This need not imply, however, that they were rude impositions upon an unwilling population. Indeed, the reverse may most often have been the case, with cathedrals (and their equivalents in other cultures) becoming representations of a proud heritage, although not
always the one that the builders intended. EMOL was the manifestation of the visions, efforts and ideas of organizations and individuals contributed over a period of almost 10 years (Weston 2001). Skills across a wide range of areas were coordinated and brought to bear and considerable obstacles were overcome, in particular with respect to rights clearance. Sites similar to it exist around the world, but the authors are unaware of any other site that combines the commitment to both high quality and high accessibility. There is little doubt that EMOL was a carefully considered project that achieved much of the considerable impact that was intended. The approach adopted to create, develop and govern EMOL has some similarities to a cathedral approach, as interpreted by Raymond (2000).

Phillips, Shephard, and Wong (2005) concluded that EMOL, employing primarily a collections-led process, left much to be desired. EMOL, after all, could only take advantage of resources and opportunities that existed at the time; it was a pioneer, preceding an era of large-scale repositories offering global e-content and universally available access devices that are now mobile (e.g. wireless laptops) as well as fixed. In terms of content provision it was inevitably tied to its own times and perceived (except by key visionaries) as subordinate to traditional libraries of mainly print materials. It succeeded in terms of building a cathedral and, indeed, in gathering a clergy, but arguably it failed in terms of gathering a congregation who would become actively involved in it. There is much in the literature to suggest that many other e-repositories currently have similar failings (see for example, Harley, Henke, and Lawrence 2006).

The e-repository as bazaar

Phillips, Shephard, and Wong (2005) suggested raising the profile of MPS objects through:

- targeting clearly defined users, with a particular emphasis on overcoming obstacles they have encountered;
- changing EMOL to streamline delivery to a wider range of users;
- transformation of the procurement process, making it active and community-based;
- supporting and rewarding teaching staff as part of a self-sustaining continuous improvement process.

The user-based collections process, described by Calverley and Shephard (2003), is inherent in the above recommendations. In some ways it parallels Raymond’s bazaar. For example, given a large enough user base of annotators and co-evaluators the user-based collections process could lead (in parallel to the bazaar outcomes) to almost every content gap being bridged and teaching and research applications becoming supported, developed and evaluated. Consider also Raymond’s pre-conditions for the bazaar, made in his publicly available audio presentation (Raymond, n.d.). These preconditions have been paraphrased below and applied to e-repositories.

- Firstly, only with an effective means of communication such as the Internet could open source processes play a role; only with the means to pass work back and forth could decentralized administrative and organizational issues be handled. In the area of e-repository development this pre-condition exists, in that most teachers, in particular in further and higher education, regularly use e-mail. However, more challenging might be whether email and online habits are sufficiently ingrained to make the bazaar happen (in the sense that software ‘hackers’ are clearly ingrained in this medium).
- Secondly, there is the plausible promise, the psychological hook that catches the interest of community members. This would act firstly to attract individuals’ interest, but more, it must go on to create the cascade, the bandwagon that forms into communities of practice (Wenger 1998). With the development of Linux this hook was an extremely simple newsgroup
posting, but it was effective and generated tremendous interest. In the case of e-repository development there is the promise of multimedia resource flexibly available 24/7. Students of the ‘net generation’ appear to have seen this promise, but arguably apply it to their social lives, rather than learning habits (Oblinger 2005; Lankshear and Knobel 2003). The e-repository bazaar for learning and teaching requires teachers to see this promise, particularly in the area of instructional design.

- Thirdly, whilst intellectual challenge is the basis of any plausible promise, more is required. For challenge to translate into time and effort a ‘gift culture’ is needed, in which giving matters and reward is intrinsic. A greater good or civic pride is needed that is shared by many. Crucially, there needs to be ‘post-scarcity economics’, so that community members are not tied to sweatshop routines to make a living. This may still exist in many areas of academia, as evidenced by a willingness to work at all hours to contribute to a community of practice, but is possibly being eroded.

- Fourthly, there is what Raymond called Torvalds’ sixth sense for good engineering. He was short on ‘design brilliance’ but he could see the ‘least effort path’ and knew how to avoid the pitfalls. On top of that was Torvalds’ skill in charming people into working with him, through a deeply honest sense of involvement. After all, there was, as Raymond (2000) put it, ‘no way to punish or coerce over TCP/IP (the internet protocol)’. Leadership is needed within the e-repository bazaar. It must include a professional honesty that others will take to, but also an intuitive grasp of both the art of the possible and art of the soluble (Medawar 1967).

E-repository development might, therefore, have much to learn from the open source movement. Phillips, Shephard, and Wong (2005) advised the JISC to seek to become a partner in the exploration of the role of digital moving pictures and sound in tertiary education and research. They offered advice on evaluation processes that will allow the academic community to assess and develop the impact of digital time-based media on learning and research. They advised the JISC not to underestimate the scale of the undertaking. Teachers in further and higher education need to be encouraged to research their learning and teaching methods. They need to be supported in their activities, and the JISC will need to understand their concerns over a range of technical and operational matters, such as the delivery of digital media, authentication processes and how resources are searched and described.

**Focus on community engagement**

Clearly the ‘collections-based’ versus ‘user-based’ paradigms for developing and governing collections of resources are different from cathedral versus bazaar comparisons for software development. However, it is possible to identify overlap between these two different paradigms and use both to search for techniques that could contribute to that crucial area of e-repository development and operation – community engagement.

At a detailed level we might be seeing this happening already. For example, the Merlot repository of learning resources (http://www.merlot.org/merlot/index.htm or) accepts a wide range of contributions from participants, but uses a range of structural elements to ensure that inclusions have at least some disciplinary acceptance and authority.

On a broader scale we see the unequivocal emergence of tools like the active world-wide web, otherwise known as Web 2.0 (O’Reilly 2005) and the concomitant explosion of information giving in addition to information retrieval. Through new technology the role of questioning, not just listening, has risen to the fore and the embryo of a possible radical shift in management thinking starts to take shape. Before the rise of such technologies and their associated communities,
generating world class products using the bazaar approach would to most people, perhaps even Linus Torvalds, have seemed absurd. Additionally, as with all aspects of civilisation, affluence enables the allocation of human resources on a scale simply impossible under subsistence conditions. New styles of management, beyond simple command and control, may now be commonplace.

In terms of e-repositories for learning and teaching, two factors of direct relevance to community engagement may be paramount.

- The strong, pre-existing traditions and values of teaching and learning within an institution and within a discipline. These are often tacit and implicit and may inevitably be preserved within a new technological environment.
- A trend to continually improve and innovate technologically, pedagogically and socially that is possibly inbuilt in the teacher’s mindset.

Laurillard (1993) interpreted such contexts with a prescriptive view, suggesting that teaching strategy should be generated through instructional design, intelligent tutoring systems, instructional psychology and phenomenography (phenomenological psychology). She suggested: ‘Each one (of these areas) provides a link between an empirical base and a principle for design’ (Laurillard 1993, 72). Salmon (2005) gave a more pragmatic view through her e-tivities five stage model of e-moderating (access and motivation, socialization, information exchange, knowledge construction and development). Inevitably, however, each provides a prescriptive, cathedral-like approach and although business management theorists (for example, Christensen and Raynor 2003) have identified how such approaches can achieve engagement and innovation, there remains much to be gained by exploring beyond.

In relation to moving pictures and sound repositories, Asensio (2003) has also commented on lack of community engagement and lack of processes to encourage it. She suggested that there are very few initiatives to promote the sharing of practice and networking and that there is a critical need to invest in community development and growth. The sense of community, theorized by Wenger (1998) and addressed by Asensio and others, is what Raymond’s bazaar illuminates. Through the analogy of the bazaar he described the characteristics of that community in a holistic way. This enabled him not only to describe the freedom and openness of that community but, crucially, the controls within it as well. For example, he described the peer pressure, the sense of belonging and the intrinsic desire to solve problems as factors that relate to control or governance as much as to routine functionality. In any functional e-repository we must equally expect some control, but perhaps not the complete absence of anarchy. The bazaar approach has led to the outstanding success of the MPS web site YouTube (Gomes 2006) that has, for example, built a strong sense of community amongst users. This is demonstrated by the volume of contributions, the comments from users and the generally strong, unanimous views expressed about authority in general, and copyright laws in particular.

However, in a search for video clips on the ‘Big Bang’ (March 2007) over 3500 hits were produced, the vast majority of which were trivial and only a tiny percentage related to The Big Bang Theory of the formation of the universe. The clips are not presented in order of date, author or sub-topic. It is not clear which of the clips appear with the permission of the copyright owners (and which may have to be taken down when the owners discover them). These anarchic aspects of YouTube may make it largely unacceptable for use as an academic resource – a video clip which a lecturer bookmarks for their students to view must remain available when the students need to view it and should not have been placed in the public domain without the permission of its rights holder. Even so, some academic users do exist (Geirland 2007) and these users must be struck by the differences in community engagement illustrated by the two e-repositories for moving pictures and sound, YouTube and EMOL.
Conclusions
This paper offers an insight into the development, use and governance of e-repositories illustrated by the bazaar and cathedral analogies and by a comparison of collection strategies that focus on content coverage or on the needs of users. It addresses, in particular, the processes that encourage and achieve community engagement. This insight is illustrated by one particular e-repository, the EMOL service. We observed that this particular repository derived from a content-driven philosophy and was managed more like a cathedral than a bazaar. As such it missed out on many attributes of user-led collections and bazaars and, in particular, had less community stakeholder engagement than was desirable. The concept of control is a significant issue, but one that can, at least by analogy, exist within a bazaar. Control must be in place within the context of education, but noting the wider envelope of interconnected constraints such as finance, technology, legality and organization. Shifts will occur within this envelope and this is where the bazaar, as described by Raymond, can have influence. For example, it will almost certainly be through the bazaar, rather than the cathedral, that the use of Creative Commons (Lessig 2004) will lead to potential solutions to problems of intellectual property rights. Similarly, it will be through the bazaars that arise through Web 2.0 applications (O’Reilly 2005) that information giving as well as information retrieval will evolve and encourage the user-based collections process envisaged by Calverley and Shephard (2003). It will be through open and trusting but at the same time professional relationships that the world class e-learning repository will arise.

References


