Catherine Candee is director of publishing and strategic initiatives in the Office of Scholarly Communication at the University of California (UC). Here she talks to Richard Poynder about UC's eScholarship Repository, and outlines her vision of the future of scholarly publishing — a world in which universities would retain ownership of their scholarly output, and make it freely available on the Web via a network of institutional repositories like the eScholarship Repository.

RP: The Office of Scholarly Communication is based in UC's California Digital Library (CDL). Are you a librarian by background?

CC: I am.

RP: What does your job entail?

CC: My mission is to marry available technologies with ongoing experiments aimed at finding new publishing models for improving the scholarly publishing system.

RP: Your job grew out of the so-called scholarly publishing crisis did it?

CC: Very much so. We faced a situation in which spiralling serials costs were literally killing the University of California. Today we spend about $27 million a year on licensed content. However, while my job certainly grew out of the scholarly publishing crisis it was also a response to the development of new technologies.

RP: You mean in the sense that the Web has allowed the University to respond to the scholarly publishing crisis in ways that would not otherwise have been possible?

CC: Exactly. In 2000, for example, we launched the eScholarship program, which was created to exploit technologies that can help us reduce the cost of scholarly materials, especially journals.

At the same time we wanted to get much closer to the classroom, and to the lab, and to find out if there were ways that the library could better support faculty — both in their teaching and in their research. As a result, we learned an awful lot about what faculty were doing with new technology, what ongoing experiments were underway (or planned), and where faculty needed help.

eScholarship Repository

RP: I guess the most visible outcome of that program has been the eScholarship Repository, which was launched in 2002?

CC: Right. And the eScholarship Repository grew organically out of our efforts to find out what faculty were doing.

RP: Talk me through the process you went through?

CC: We discovered that since the structure of the literature varies considerably from discipline to
discipline, faculty were taking advantage of new technology in vastly different ways.

We found ourselves, for instance, trying to support people in humanities who were using GIS systems with a temporal aspect in order to publish their data in new ways; we found ourselves helping the social scientists put working papers online; and we found ourselves helping people in performance arts who wanted to capture performances and store them.

Before long we were involved in twelve different projects, and as we started trying to develop the infrastructure to support all these projects we found ourselves being pulled in so many different directions that we realised that none of them would be sustainable if we tried to support them all. We also knew we would need to support further projects as they arose.

So we realised that if we were going to meet all these needs adequately we would have to find some sort of generic solution. And since the predominance of the materials that we were supporting were textual we began — in a very pragmatic way — looking around and testing repository software.

RP: So while the roots of the eScholarship program lay in UC's attempt to address the problem of journal price inflation, the eScholarship Repository grew out of your growing knowledge of the specific needs of faculty?

CC: Exactly. The journal pricing issue drove us in the library to seek new solutions; but it wasn’t journal pricing that drove faculty to try new things. In the end, therefore, the eScholarship Repository grew out of the opportunistic use of new technology by faculty, and the decision by UC library to establish new ways of helping faculty.

RP: What was the appeal to faculty of your offer of help?

CC: The attraction for most faculty units was that they had a lot of materials that were at risk: they were putting seminar papers up on web sites that were disappearing, and they were frustrated at trying to manage huge inboxes — because in the preprint environment there were manuscripts flying around in a generally unmanaged way, and they were not being properly preserved.

RP: Initially you built the eScholarship Repository with the EPrints software, which was developed at Southampton University in the UK?

CC: Right. We started with Eprints, and the aim was to create what people now call an institutional repository — a repository where faculty could put materials (text and images) that they wanted to disseminate, or actually publish.

A different model

RP: You later switched to the bepress software. Why?

CC: We found it so, so, so difficult to get faculty even to test the EPrints software that we abandoned the idea of providing a platform for faculty to individually publish their own works. Around the same time we serendipitously encountered the bepress software, and right away we could see that it would to allow us to do something much more important.

RP: How do you mean?
CC: We could see that if we used the bepress software the repository could also support peer-reviewed publications. Consequently, by the time we launched we had switched to a different model, and we had adopted the bepress software.

RP: How was the model different?

CC: The bepress software allowed different units within the University of California to become the gatekeepers, with all the editorial and administrative ability resting with an academic department, an institute, or a research unit, rather than with individual faculty, or with the library.

RP: So where EPrints software assumed that researchers would do the inputting of papers themselves, bepress software was more suited to third-parties depositing them?

CC: That is one difference — although, because the software is difficult to use, Eprints submissions are often managed centrally. Additionally, the bepress software lent itself to the size of UC; and it allowed the University to decide exactly what it wanted to put in, and to brand everything in the way it wished.

RP: You were also able to outsource the hosting of the eScholarship Repository to bepress?

CC: Yes. It is hosted by bepress, but preserved here at CDL.

Harvesting citations

RP: You said the original aim was to create an institutional repository. There is some debate today over the terminology used when talking about repositories. Would you say that the eScholarship Repository, as it has developed, is still an institutional repository?

CC: I would. It is hosted by this institution, and it is managed by this institution.

RP: Certainly most agree that an important role of an institutional repository is to allow a university to make its peer-reviewed papers freely available on the Web, and thus "open access". As it happens, UC is being very proactive in this regard. It has, for instance, introduced a metadata harvesting program designed to track down papers published by faculty, and it then asks the authors to deposit postprints of those papers in the eScholarship repository. How does the program work?

CC: What we are doing is harvesting citations. We then send them to faculty members saying that the listed works may be eligible for inclusion in the eScholarship repository. It is a way to alert them to the repository, and to the fact that they have content that could be placed in it.

RP: How do they then deposit the postprint?

CC: The message sent to faculty is clickable, and when they click on the link it brings them directly into the repository, where the citation data for the paper automatically fills out the repository metadata fields for them. This, by the way, is the one case where we allow authors to put their content in directly themselves.

However, we also allow them to use a proxy — so they can legally assign someone else to put their papers in for them. The aim is to make the process as easy as possible, because time is the biggest constraint when it comes to getting faculty to participate.
Rights clearance

RP: And you have contracted bepress to do rights clearance on the papers?

CC: Right. After the papers are submitted we pay bepress to check the rights on them. That was a concession to the fact that bepress’ business would be threatened if they got sued for allowing something illegal to be put into the repository. This part of the process is both onerous and expensive, and we hope we will not need to do it at some point in the future.

RP: I'm told you have acquired about 1,000 papers in this way. Does that figure represent the total number of postprints in the eScholarship repository?

CC: Yes, that is the total number of postprints in the eScholarship repository, and around one tenth of the papers currently available. [At the time of the interview there were 10,373 papers in the repository].

RP: How do the other papers get into the repository?

CC: It depends what kind of paper they are. The working papers, the technical reports, the state reports, and other professional materials that have not been peer reviewed are all deposited into the repository by the units, who act on behalf of the member of faculty.

RP: So the units allocate a member of staff to do the inputting?

CC: Exactly. And with the seminar papers we allow faculty to create the metadata for papers that are going to be given, and to add a sort of place holder in the calendar. Then, later, someone goes in and adds the paper. That is still the responsibility of the unit however.

Mandating researchers

RP: 1,000 postprints is a small drop in the ocean I guess. How many researchers are there within the UC system?

CC: UC is the largest public research university I know of. It has ten campuses and around 16,000 faculty and researchers.

RP: When you ask faculty for a postprint is it a request or a demand?

CC: It is not a demand. Clearly, incentive is the single biggest issue for getting content in. Awareness is another issue, so we are just starting some market research to discover what percentage of UC faculty even know about the repository. I suspect it is less than half.

RP: So you still have work to do in publicising the repository?

CC: We do. While we are very excited that we have more than 200 departments participating in the repository we have no idea what percentage of the faculty know about it; and we have no idea what percentage would participate if they did know — because there is no overriding incentive for them to do so today. We need to understand the situation.

RP: As your experience shows, creating a repository is only half the task. You then have to fill it.
For that reason there are growing calls for funders to mandate researchers to self-archive their papers. Do you think that that is the best way of filling institutional repositories?

CC: Well, I wouldn’t say that our purpose is simply to fill institutional repositories. We built an institutional repository as one way of providing an alternative to the current publishing system, and to give faculty something to do with that copyrighted material that we keep saying shouldn’t be given away to publishers.

It may turn out that institutional repositories aren’t the way to go however. For that reason we are also interested in encouraging faculty to manage their copyrights differently, and to consider who they give their manuscripts to, and where they commit their editing and reviewing time. So our main focus is in accomplishing that, rather than filling repositories.

Managing rights

RP: Do you nevertheless anticipate that funders will eventually introduce mandates?

CC: Actually we expect that universities will make some sort of a mandate before funding agencies do. In this regard there are a number of white papers floating around the University of California right now. We are waiting to see what happens to those.

RP: Yes I saw that. Given what you say about rights, I’d be interested to hear more about the Scholarly Work Copyright Rights Policy white paper. This proposes that UC faculty "routinely grant to The Regents of the University of California a limited, irrevocable, perpetual, worldwide, non-exclusive licence to place the faculty member's scholarly work in a non-commercial open-access online repository." Would this apply only to journal articles or all the works of faculty, including books?

CC: Ultimately it is intended to apply to all works, but starting with journal articles.

RP: The aim is to take the citation harvesting program one step further is it?

CC: Actually the white papers are an initiative of faculty, not us. Indeed, the most exciting aspect of it is that these papers were put before the Academic Council by the Scholarly Communication Sub-Committee of the Senate. This is significant because, as everybody knows, faculty are far more likely to listen to faculty. But if implemented the white papers would clearly allow the University of California to significantly extend the work we have been doing with the eScholarship Repository.

RP: Who will decide whether the white papers are accepted?

CC: I haven’t followed a white paper process like this before so I don’t know exactly. But the papers were immediately approved for circulation by the Academic Council, and they’re being circulated and talked about as we speak. I imagine each of the academic senates on the campuses will take them up, discuss them, and they will then be brought back to the system-wide senate.

We can’t know what the outcome will be but, at a minimum, there will be an awful lot of consciousness raising. And who knows, perhaps UC faculty will indeed choose to act in concert to change the way that they manage their rights.

RP: What is the likely timing for a decision?
CC: As I understand it, the aim is to get things passed and through the system before next fall.

RP: If it does go ahead would you envisage a postprint mandate following behind it?

CC: Yes.

RP: And you would welcome that?

CC: I would. While I don’t find the postprint issue as interesting or exciting as trying to encourage new forms of communication, it is strategically important — because it would allow us to put in place a production-level service capable of managing UC copyrighted material, which would better prepare us for the future.

Stepping stone

RP: So while your long-term ambitions are greater, you are keen to see UC researchers deposit their postprints in the eScholarship Repository as a matter of course?

CC: That's right. We now view the postprints project as a kind of stepping stone, or a means to an end, to changing the paradigm, and of educating faculty. Part of that process means getting hold of a lot of content so that people can see there is real value in having this managed by the University, that there is a value in making it open access, that there is value in being able to speed up communication, and that there is value in having more direct control over it. So we are keen to see any initiative that will drive content into repositories and help change the way people do things.

RP: You expect to see universities taking a much greater role in publishing in the future?

CC: I do. And as the number of repositories being developed around the world has taken off so more and more people are beginning to see them as publication and communication tools.

RP: Indeed, UC has already started publishing electronic open access journals in the eScholarship Repository hasn’t it?

CC: Yes. We have used the platform for all levels of peer review. I should add, however, that we weren’t aiming specifically to get faculty to launch new academic journals. There are too many journals already!

But we recognised that there were some niche journals that could use the repository software well. There were, for example, some in-house UC magazines that have been around for a long time that really just needed hosting. In fact, although we don’t think that it is going to be the primary use of the repository, there is now a queue of things moving into the journal part of our publishing section.

Alternative publishing models

RP: Can you say more about the alternative publishing models you envisage?

CC: Sure. One of the things that is so interesting about the repository is that about half the materials are never published in the traditional sense: they have a life all of their own. So, for
instance, many conference papers and seminar papers that would never be published are now able to live as part of the Web, and of the scholarly record.

**RP:** This is like the concept of the *Long Tail*?

**CC:** That’s right. There is one other really interesting initiative that I should mention. You maybe know that we have something called eScholarship Editions?

**RP:** No. Tell me about the initiative?

**CC:** Actually we don’t use the repository for this: it runs in-house on a home-grown open source system we call XTF — or eXtensible text framework.

Anyway, eScholarship Editions are scholarly monographs encoded in XML. And we have marked up the backlist for University of California Press titles and are now trying to redefine a workflow that would allow us to publish in XML going forward in a cheaper way.

**RP:** eBooks?

**CC:** Yes, eBooks. And in the meantime — because they are so expensive to produce — we have also started using the eScholarship repository to do monographic publishing. We have four monographic series underway so far.

**RP:** What is the rationale here?

**CC:** As you know, the corollary to the serials crisis is that libraries have less money to buy monographs, and so fewer monographs are being published. The fact is, however, that an awful lot of monographs could be published if the UC Press had more editorial bandwidth.

So we have been experimenting with empowering UC Press editorial boards, or faculty editorial boards, to become, essentially, publishers. In this way we can extend the work of UC Press.

**RP:** The eScholarship Repository has also effectively become a publishing platform for the University of California Press then?

**CC:** Exactly. It won't be used for most of the monographs: there will be a tiered arrangement. So there will be some fabulous critical editions and glorious books that won’t ever go through the repository, but there is an awful lot of really good material out there that is worthy of publishing, and that faculty very much want to publish. We are happy to share the load in order to get it published.

**RP:** Will these monographs be published in print or electronic format?

**CC:** They will be printed if there is sufficient demand for them. The wonderful thing about publishing the first copy digitally is that print publication can be taken up by UC Press as a separate, business decision. Ultimately, we hope to offer the monographs as Print-on-Demand.

**Something narrower**

**RP:** It's clear you have a very broad view of the role of an institutional repository. Advocates of self-archiving, by contrast, insist than an institutional repository should only ever be viewed as a
postprint archive. What's your response to that view?

CC: I think it is unfortunate that the term institutional repository has come to mean something narrower. As I say, the postprint component is the least interesting and ultimately the least important part of this. So while right now it is tactically extremely important to deposit postprints, ultimately I envision a very different arrangement between universities and publishers than we have now.

RP: You believe universities should be in control of the publishing process, rather than managing papers that have been published by someone else?

CC: That's right. Eventually I hope all the content will be hosted and managed by universities themselves, and the publishing services would be in the form of added value. So, for instance, a published article would refer back to the raw article in the repository.

RP: What sort of added value services do you envisage?

CC: In addition to peer-review, I can see scope for various kinds of indexing services, and for aggregation services. People love to go to processed stuff that offers them a limited view of the content. So there could be discovery mechanisms to allow people to do research in their discipline regardless of the location of all the materials — this could be a service, for instance, that harvested content from a bunch of institutional repositories and then sorted and metadata enhanced that content along a discipline line. That is certainly a model we are looking at.

RP: On the other hand, there is a school of thought that argues that rather than posting papers in institutional repositories, researchers should post them directly in disciplinary repositories — using, for instance, subject-based services like arXiv or PubMed Central.

CC: Until I realised how differently the disciplines were using technologies I also thought that discipline-based archives were the way to go. I don't any more. Moreover, now that there is a fairly well developed layer of institutional repositories becoming available it makes more sense for a service layer to develop on top of those repositories for researchers who want to distinguish material on a discipline basis.

But to go back to your earlier question, I believe it would be just too bad to limit our vision of the institutional repository to postprints alone, and to not exploit their potential for enabling faculty to put all kinds of creative output in them.

Deterrent

RP: What worries self-archiving advocates about this is that if universities try to make institutional repositories too broad in functionality they could delay the transition to an open access environment; that we need to stay focused on the narrower view until OA is achieved. You are arguing that we need to plan for the longer-term future from day one are you?

CC: I think so. Moreover, I don’t see why a broader view would slow OA down. It is a matter of getting the right platform and getting things moving so that faculty can see that there are other things that can be done.

RP: One obvious speed bump, perhaps, is the cost of building a repository. While it only costs a few hundred dollars to set up an Eprints server, the kind of repository you are building is
inevitably far more expensive. Indeed, in a recent issue of the INASP newsletter, Ann Okerson, a librarian at Yale, estimated three-year start-up costs for hardware and software alone for such a repository at over $300,000. While that might be fine for a UC, a Yale, or a Harvard, it will surely act as a powerful deterrent to any smaller institution considering setting up an institutional repository?

CC: Sure. In fact I doubt every single school will have its own institutional repository. More likely their content will be hosted by the larger schools like ours. We are in conversations right now, for instance, with California State Universities — trying to figure out how to partner to make it easier for CSUs to start out.

RP: So some of the costs could be shared between institutions?

CC: Possibly. Certainly we are open to partnerships, and to maybe extending services to other universities. I should stress, however, that we are not open for business yet. We are only exploring possibilities.

Natural caretakers?

RP: Are librarians the natural caretakers of institutional repositories?

CC: Well, there is a strong argument for saying so. Libraries have a good history of looking after things. At the same time, however, they have not historically been identified as publishers. So to the extent that this is a publishing service I can see that it could work against a repository to be over identified with the library. The best and the healthiest approach would be an alliance between faculty and libraries.

RP: I wonder if we might see increasing tension between researchers and librarians over the issue of institutional repositories? I ask because the primary aim of researchers is to achieve maximum impact for their research; librarians, by contrast, are looking to create large digital libraries or even, as in the case of UC, complete publishing systems. Could this threaten the historic relationship between librarians and researchers?

CC: I can see such a tension theoretically: where resources were limited, for instance, the aim of building a digital library could seem to stand in the way of getting publishing out quickly. But ultimately I think you are presenting a false dichotomy.

RP: If someone from another university was looking to create an institutional repository and asked you for advice how would you reply?

CC: I would tell them to consider all of the new forms of communication that are taking place and to work very, very closely with faculty to determine what they need.

RP: Looking to the future, how important do you think institutional repositories will prove to be in the scholarly publishing process and will they be seen as an alternative to the traditional system or as an adjunct?

CC: In the short term I think they will be quite important. I don’t see them as a replacement but, as I mentioned, I really think we are heading towards a layering of services, where an awful lot of raw content will be managed more responsibly by universities, and publishers and aggregators will develop all kinds of services to add value to that content.
**RP:** So the basic content would always be free and openly available in an institutional repository, but those who wanted to could go to a publisher and pay to use enhanced search and aggregation services?

**CC:** I think that's right.

**RP:** OK. Thank you for your time.