AN APPROACH TO MANAGING REPURPOSING OF DIGITISED KNOWLEDGE ASSETS

Diarmuid Pigott
School of Information Technology
Division of Business, Information Technology and Law
Murdoch University
Murdoch, WA 6150

Valerie Hobbs
School of Information Technology
Division of Business, Information Technology and Law
Murdoch University
Murdoch, WA 6150

John Gammack
School of Management
Griffith University10
Nathan, QLD 4111

"The question is,' said Alice, "whether you CAN make words mean so many different things."
"The question is,' said Humpty Dumpty, "which is to be master – that's all."

ABSTRACT

Managing digital assets not only involves internal distribution, accessibilities and efficiencies, but also wider protections involved when derivative works or repurposings result from external distributions. The inherent polysemy of media artefacts has consequences for the reuse of such artefacts in multimedia systems. The multiplicity of possible meanings of a single media artefact depends on context and intention: reuse of a media artefact thus almost inevitably leads to the artefact being altered to suit the new context, and potentially losing its original meaning in the process. A system of rules is required to identify the circumstances under which a media artefact may or may not be reused throughout its life. Within this context of practice we detail a protocol of usage status and reuse request. In this protocol, usage status is assigned to a media artefact by its initial custodian and the outcome of any reuse request is determined by the intention of the reuse, the current usage status and the implied new usage status. Possible outcomes are forbidding change, permitting change, artefact forking, and changing artefact custodianship. We present a series of examples to illustrate the range of possible scenarios and their outcomes.

INTRODUCTION

All media artefacts in use represent points of information. Any use represents an organisational context, regardless of whether the media artefacts are shared or not, and any shift in this context represents a shift in meaning: when an artefact is reused or repurposed it acquires a diffuse or altered informational state which may work against the original intention that existed at its creation or adoption. Although this problem can logically occur when there is only one user, the owner, it is manifest mostly when artefacts are shared. This problem becomes more than academic in the context of the development of organisational media capital, as the considerable investment in the creation of media artefacts can be a recoverable resource through planned reuse.

Organisations large and small are increasingly critically dependent on their information base and the knowledge resources that deploy information intelligently. Their customers and markets are being driven by Internet enabled technologies, and their products and services mediated electronically. For organisations of even moderate size, digitised multimedia artefacts, customer records, product designs, transaction histories, business processes and methods comprise valuable intellectual property impacting on operational and strategic decisions, and conferring comparative advantage internally or competitively. Storing, accessing and securing, querying, and analysing this resource are already essential activities, with increased potential for effective operation enabled through useable web-based interfaces and architectures, and accessible multimedia repositories.

Beyond these proven systems for information management, however, are advanced technologies required for managing the *knowledge* that can identify and apply information in current contexts, and that can flexibly recombine components of information and other resources to provide new products and services, customised intelligently and market responsive. The ability to add value to an information repository through intelligent management and reuse of the knowledge involved in so doing gives an organisation in any sector a unique edge when the base technologies themselves are not a differentiator of competition.

[10] From January 2002. The work was done when the third author was at Murdoch University.

The inherent problem of media artefacts as information points is that they are polysemous and not at all self-descriptive, so that some regularised form of data must be stored with them to permit conveying of meaning, storing, sorting and contextualising. By convention this is called metadata, but giving the practice a label only defers the problem, as the formalisation of rules of storage merely imports all of the problems that have been the subject of study of information organisation for centuries.

Much tacit knowledge, providing the interpretive context within the minds of knowers processing information, can be elicited and made explicit through various techniques (Gammack, 1987), including unobtrusive modelling of usage patterns (Gammack, 1991). Knowledge is also shared, and it is this collective knowledge that typically allows an organisation to identify relevant data, to learn and to leverage its information resources effectively. Its independent representation has however proved an enduring research problem. The issue of including *context* in knowledge codification and storage is a great challenge in this field and one of the important research questions identified by Alavi and Leidner (2001 p128).

Producers of, and investors in, recorded digitised content typically have the right to determine the legitimate contexts of use, and the moral indenture associated with reuse or repurposing. Internet technologies enable the reproduction, reformatting, selection and derivative product development from internal repositories of material, but also potentially by third parties in ways which may compromise the integrity of the source material. Already, amateur publications and samizdat[11] academic materials of various kinds are spreading through peer-to-peer distribution arrangements, adapted, recontextualised and reinterpreted. The potential for "Chinese whispers" (e.g. Livemann, 2000) to occur is obvious, but even within corporate intranets, restructurings, mergers and staff turnover means that recorded corporate knowledge can mislead, both internally and externally.

For effective reuse to occur, relevant contextual information accompanying the creation of digitised artefacts must be recorded. Such information qualifies and constrains the possibilities of misrepresentation and misinterpretation, whilst admitting a variety of legitimate uses. Issues include identifying provenance and legitimate ownership of digitised artefacts, allowed uses under different representational and contextualisation protocols, and the potential for reformatting and repurposing of material in knowledge repositories, using standard metadata storage mechanisms.

Apart from emerging delivery platforms and media formats, the distributed and global nature of the Internet, encompassing numerous communities, jurisdictions and accepted standards further complicates this situation. A current case reported in the Sydney Morning Herald typifies the issue (AAP, 2001). Here a text originally produced for consumption within one physical community (the USA) is considered (post-Internet publication) to defame its subject in the community where he is best known (Victoria, Australia). Whilst legal rulings and precedents will be set and established in due course, the principle of community context for information content is the point of law of relevance here.

We recognise that media artefacts, and symbolic materials generally, are open to selective and contextual interpretation by others upon publication. This however does not exclude the intentions of the original author, nor the community context within which the work was produced and to which it references. Without appeal to philosophical conceptions of knowledge admitting the transcendental, the universal or the transcultural, the frame of reference for any symbolised artefact must ultimately be embodied in community and cultural norms, at the level at which laws are shaped and socially agreed. This is important, since it conditions the status of an artefact and its potential for legitimate use.

To manage the use of digital assets involves addressing issues of context, ownership and provenance, going beyond mere access controls and document management. Instead the deep semantics associated with usage of particular digitised artefacts must be represented to a degree sufficient to avoid intentional or unintentional misrepresentation and misuse. The legitimate categories of use may be revised or redetermined by the asset owners or relevant communities of practice, but effective mechanisms to allow this are required if these artefacts are to retain their full authority as knowledge sources, and their provenance to be established.

---

[11]Samizdat (from the Russian for self-publishing) refers to the clandestine network of reproducing and distributing unauthorised literature in Soviet Russia, and now by extension a generic term for self-publication which bypasses conventional production and distribution channels (such as Internet sites and Usenet)

RECONTEXTUALISATION, SEMANTIC DRIFT AND THE PROBLEM OF MONODIRECTIONALITY

The potential for recontextualisation and either accidental or deliberate out-of-context repurposing is a particular instance of the generalised phenomenon of *semantic drift*: the gradual change in meaning that can happen to an artefact during its lifetime.

Consider an old photograph album found in a second hand shop. Inside is an ordered sequence of brown photographs held on the pages by white paper corners. If the owner of the album has been diligent, then under each photograph in white ink will be a description of the occasion, time, date, place and subject matter. The point of keeping photographs in an album is partly convenience of storage, but mainly preservation of context. This is such an efficient mechanism of preservation that 60 years on intimate details about the lives of people recorded in photographs are easily retrievable and comprehensible, even though the people are unknown to us. We would know, for example, the names of a family group on the beach, and the identities of each one of them from the inscription underneath.

The same photograph in serialised stock of a seaside photographer would lose its personalised context but acquire a new context in the record of the photographer's professional life. Within this context it is almost impossible to recontextualise the identity of (e.g.) the photograph of the family on the beach. With further reuse the photograph acquires a new context of display and retrieval: if the album or the stack becomes part of a museum collection, the photograph may become something that stands for "seaside holidays of the forties" when it is part of a historical display – the sense of personal identity in the photograph is lost.

Artefacts in use are in use precisely because one can ascribe intentionality to them. This is the difference between finding a photograph in a junk shop and being given one by someone who describes it to you. With the former, while the observer can infer a potential meaning and significance, there can be no certainty to the inference. With the latter (barring deceit and inaccuracy) he or she is provided with the photograph's meaning and significance. In any information system that uses media artefacts, this meaning and significance must be present before the artefacts can count as points of information.

A useful analogy here compares the archaeological analysis of cultural artefacts with the sociological: the sociologist can proceed from the known to the unknown, asking questions of usage from the current owners, and building up a meaning-and-significance picture which can be verified with those owners. The archaeologist, on the other hand, is in the invidious position of proceeding in the opposite direction – from shards, broken cookery vessels, hearths and inscriptions back to a living culture: trying to infer the context from the object normally found within that context. This type of reverse engineering works with a footprint, but not with a cultural artefact. The fact that we bother trying to do this in archaeology is because there is no other means at our disposal of working out the context. And we bother precisely because they are artefacts – they must fit within a context somehow, because they were made within a context.

Media artefacts in an information system, then, can only be given their meaning and significance by accompanying metadata. By storing meaning-and-significance in metadata, we can always get back to a knowledge of the original context, even if all we have left of that context is the metadata.

But where metadata must be created is at the point of the artefact's creation (or accession), and this leads to two separate problems. Firstly, there is a distinction to be made between the metadata that is actually stored, and the potential set of all metadata that could possibly be stored for any artefact or class of artefacts. This potentiality can exist either as categories of metadata or as choices to be made within a category of metadata.

Secondly, and more importantly for our purposes here, the subsequent reuse or repurposing must involve either ignoring the new metadata, or overwriting or augmenting the original set. Overwriting or ignoring metadata leads to a loss of information, while augmenting metadata may render it meaningless (e.g. a picture being both "attractive" and "ugly" from different points of view).

The solution logically would lie in a system that permits the storage of usage and context in a continuing historical metadata narrative. This could be an MPEG-7 compliant description (ISO, 2000) which also features versioning information, or in a set of XML RDDL tags (RDDL, 2001), or simply as transactions within a regular DBMS.

The pre-eminent means of displaying and retrieving information has become the Internet web-page browser, as the lingua franca of HTML markup and inline image display has swept other presentation media before it. The resource sharing systems based on the HTTP standards wherein images and metadata are served from a daemon via TCP have overcome many of the problems that beset universal presentation formats before now. However, this greatest common multiple solution has brought with it some profound problems for the implementation of an artefact-sharing system that involves knowledge of how each artefact is being used.

The HTML linking system is ultimately derived from the practice of scholarly citation, where one paper or monograph refers to another by means of attribution, according to certain bibliographic standards. Every paper contains a list of all (admitted) influences and sources, and so a backwards chain of referencing is possible that can extend for-

ever. However, the reverse is not true – there is no means of automatically proceeding from an article or paper to those that cite it. In academic research this service is provided by the citation indexes, which use brute-force methods by indexing all references in the main academic journals. But there can be no guarantee of completeness in this direction.

Even if the citation indexes were sufficient for making a forward chain of referencing, they would still be insufficient for retrieving the *context* of citation. This can be clarified by using the terminology of Polanyi (1966): when one goes from the citing to the cited, one proceeds from a context, one is *attending from* the citing article and is *attending to* the cited. When one goes from cited to citer, the reverse will never be true. The following axes of information and context are missing:

Generality/specificity: there is no way of knowing whether the referend is to a part or the whole of the referred object.

Essentiality/substitutability: there is no way of knowing whether only the referred object will suffice for the reference, or whether any similar referend will do.

Agreement/disagreement (or dismissal): there is no way of knowing whether or not the referring item is positive or negative about the referend.

Research/contingency: there is no way of knowing whether the referend was discovered after much research, and so has a logical trail leading to it, or whether it was discovered by chance and has less of a logical justification.

Fair-dealing and appropriateness/unfair dealing and out-of-context quoting: there is no way of knowing from the referend the extent to which the referring item is dealing fairly with the referend – quoting in context or out-of-context, using a portion that is an indirect quote of a third item, or even using ellipsis to misquote completely.

Recontextualisation in repurposing: there is only a pious hope that the recontextualising in all of its considerations is such that the original meaning is not violated.

Secondary citation: the problem of the cascading referentiality that leads to semantic drift. Here is the most significant for our current consideration: there is actually no way of determining whether or not the continuous referring and re-referring to the referend will lead to any or all of these problems to arise.

Thus we see that if the citation resides in the citer, with no reverse linking in place, there can be no possible way of reverse creation of the context of citation. These problems do not arise accidentally from careless usage: they are simply not determinable from the circumstances, as the *fact* of use does not indicate anything about the *nature* of use. Fundamental with the problem of HREF, and its replacement, the XLINK instruction, is the problem of monodirectionality.

This is in stark contrast with the original conception of hypertextuality, wherein the mechanism (and thus the process) of hyperlinking was duplex. This is what Ted Nelson envisaged in the Xanadu project (Nelson, 1990). "We long ago foresaw the problems of one-way links, links that break (no guaranteed long-term publishing), no way to publish comments, no version management, no rights management. All these were built into the Xanadu design" (Nelson, 1998).

Also missing from the property of linking as implemented as in HTTP is the point of multiple departure, inherent in all of the semantic usages of linking. This was also an area of criticism in early comments on the HTTP specification, from the hypertextuality development community (see Berners-Lee, 2000). This in many ways shows the origin of the HTTP standard in the academic bibliographic tradition (Burnard, 1991), wherein reference is a one-way mechanism, reinforced by the citation bibliographic tools that are the standard mechanism for academic research.

The difference between true hypertextual structures and convenient computational structures for web maintenance is outlined by Bernstein (2001). Links in true hypertext create structure for the reader, using recurrence to reinforce meaning. "Recurrence is the main way that people perceive a hypertext structure, the way they learn what contours they may follow and how those contours may change as the document evolves" (Bernstein, 2001).

These formalisms give semantic limitations in usage and distribution of digital assets since there is no inherent provenance or context represented in the resource, and, as Nelson indicates, neither version nor rights management is inherent.

Even with a process of reverse linking in place, however, we are still moving from the contextualised towards the context, which is impossible in terms of tacit knowledge management theory (Gammack, 1987). The existence of a formalised but unqualified system of reverse citation would still not permit the context to be recreated as the action of reverse-linking is still merely indexical, not actually the context itself. The problem of secondary citation or repurposing in particular cannot be guaranteed to avoid the problems described above: the context of citation must be stored in metadata with the link itself.

As we have seen, with the repurposing of a media artefact comes responsibility to not misuse it, not to breach copyright, or to take an image and use it in an ironic or offensive way. What is need is a statement of *purpose* by the new user, and a guarantee that the image will be used in an appropriate new manner. Even then, the owner and original user must explicitly accept the new use, to ensure that no semantic drift occurs.

CURRENT INITIATIVES FOR REFERENCE AND USAGE

Within the international information community there has been an intense period of research and development in both the systematisation of metadata, and the establishment of conventions for content-in-creation standards for assigning metadata to media artefacts.

A large amount of the relevant work comes out of the digital libraries and digital media preservation communities, as it is here that change in the long term is seen as inevitable (largely because it has been met with already in the brief life of the electronic artefact), and therefore systems must be set in place now to ensure the long term survival of the identity and content of the media artefact.

The Library of Congress has identified a set of relevant challenges to constructing digital libraries (Library of Congress, 1998), but without addressing the issue of knowledge based recontextualisation of categories. Their list includes effective access to information within copyright and other legal concerns, retrieval of information across heterogeneous database sources and data formats and metadata schemes, and the construction, maintenance and enhancement of catalogued resources. Several relevant major current projects internationally are sponsored through the NSF's Digital Libraries Initiative – Phase 2 (Digital Libraries Initiative, 2001). Information-retrieval oriented work (e.g. Chen, 2000) shows one typical approach, using computationally intensive fuzzy classification algorithms applied to large collections to replace knowledge structures derived from human intermediaries. These approaches are essentially brute force and historical, and have no basis in the deeper semantic codes of human usage.

Standardised protocols are already in use in many projects, but there are ongoing efforts at improving them (McCallum, 2000). For example, the ISO Z39.50 standard specifies a protocol for searching and retrieving information from remote databases, but has a number of perceived limitations (Lynch, 1996). Some multimedia formats may prove unsuitable for long term indexing as new standards replace existing ones for representation of content. In addition some resources may be in physical or non-digitised formats.

The MAKERS system of Chen (1997) proposes an object-oriented knowledge representation, which builds on the Dublin Core (DCMI, 2001), and has the advantage of user centred extensibility, where additional fields relevant to user communities are represented without interference with permanent library semantics. Chen's proposal does not appear to have been implemented to date, and does not specifically address the problem of concurrent mutually-antagonistic metadata needs, nor the problem of potential metadata overwriting.

Also of interest is the concept of data provenance, as described by Buneman (1999) in a Digital Libraries Initiative project, since media artefacts may be extracted from one context and used in another, losing meaning or misrepresenting intention. Usage of media as a meaningful action is irreducibly based in human semantic spaces and is not considered simply a property derivable from disembodied associations (Clancey, 1997), and invokes notions of classificatory knowledge being grounded in processes of discourse (Star & Ruhleder, 1996).

In practice, the organisation of information in digital repositories does not need to be bound by the metaphors guiding the arrangement of physical libraries, enabling new design metaphors to guide the dynamic structuring of resources. Svenonius (2001) has identified key principles and functional objectives for organising intellectual resources, which point to a new type of cataloguing solution: one enabled by current technologies that addresses the traditional difficulty of constructing appropriate classifications of materials from limited category labels.

Besser (2000) cites the 'custodial problem' for digital longevity - as yet there are no traditions for preserving digital materials as there are for non-digital, and a major problem is determining who should be responsible for maintaining material in an electronic form. Besser also highlights the inevitability of mutation over time when files are repeatedly copied to new strata. He also cites the 'translation problem' wherein the change of the form of digital content often serves to change part of the meaning. This process is part of the general problem of changing meaning over time that we have termed semantic drift above.

In a recent white paper, the OCLC/RLG Preservation Metadata Working Group (2001) reviewed the state of the art in preservation metadata, using the OASIS Information Model as framework for metadata requirements for digital preservation. This model includes both the data object itself (physical or digital), and a knowledge base that is associated with the users and necessary to interpret the data object. The knowledge base itself is considered to be external to the archive, and is not maintained as part of the archival function.

Another source of metadata usage is the metadata-in-creation community, where systems such as the Digital Object Identifier (which continues the traditional cataloguing-in-publication into the digital domain) (International DOI Foundation, 2001) and the Resource-Based View work (Habann, 2000) focus on the maximisation of the information potential (including recording of context) of a media artefact by imposing metadata standards at the moment of its creation. This work does not specify the usage system, but prefers to have an ecumenical approach to the systems of delivery. However, it does not permit the management of change in the semantic nature of the artefact, as the metadata is carried with the artefact while referencing standards, and is static. Each usage is therefore independent of others, linked only by the common usage.

The conceptual schema of a hybrid emergent/structure-based system for establishing semantic links between media artefacts, and amongst information repositories generally, has been identified as a key component of rational knowledge representations by Sowa (2000). Drawing heavily on Sowa and the earlier Cyc work by Lenat (Guha, Lenat, Pittman, Pratt, & Shepherd, 1990) is the Internet Topic-Mapping initiative (ISO/IEC, 1999), which aims to provide a unified representation of knowledge in a manner that relates to its embodiment in information resources. The proposed standard has XML-encoded structures that permit the creation of loosely established late-binding semantic maps, which in turn make use of the existing hypermedia protocols and formats of the web. Gärdenfors (2000) has elaborated the case for a framework for representing information at the conceptual level, which addresses some of the difficulties with purely symbolic approaches to concept representation.

Lagoze (2000) points out that while traditional library materials have always had to deal with a complex web of relationships among information entities such as translations, editions and transcriptions, networked digital information has greatly increased the complexity of those relationships. He proposes the role of catalogue as mediator of a distributed system: "Rather than absorbing semantics (and descriptions) from distributed communities, libraries should promote the catalogue as a mapping, or interoperability mechanism, amongst distributed descriptions. Technologies such as RDF [W3C, 2000] and its schema language... make it possible to undertake such a mapping role amongst individual descriptions that are distributed across the Web". Such a system would require protocols in place to ensure against semantic or physical alteration, and the problem would increase, rather than decrease, with the scale of usage.

Bearman and Sochats (1996) identify situation variables – "inflation, changes in permissions based on elapse of time since the event, re-engineered business processes, etc." – that in alteration can create a change in the meaning of a digital asset. Semantic consistency could be created by a stored object that is "encapsulated by metadata necessary to ensure its integrity and longevity", and he proposes a record keeping system or warehouse where it would be kept intact. While this system is a useful model, it is geared more towards maintaining accuracy in a single meaning, rather than permitting the maintenance of concurrent variance in meaning and ensuring the consistency of those multiple meanings through time.

Finally, the LOCKSS model proposed by Reich & Rosenthal (2000) to preserve access to scientific journals published on the Web, works by using redundancy by permitting multiple independent copies to be maintained in different locations, with integrity ensured by a system of digital signatures. However while this method preserves the content of the article, it is at the expense of a central reference scheme.

We can see that there are two components missing from these current initiatives for meaningful preservation of media artefacts: the cumulative recording of the context of usage (which enables semantic drift to be recorded) and the means of censure/restriction to permit/restrict untoward usage.

We now turn to the consideration of what procedures could be put into place in a media artefact repository to enable these requirements to be met, and present a proposal for a protocol to manage the reuse of digital assets.

A PROTOCOL FOR REUSE

Let us consider the acquisition and subsequent reuse of digitised photographic images in a museum archive. As we have discussed, when an image is added to the archive, it is essential that metadata describing semantic purposing is included in order that the base requirements of preservation of context be met, to manage any potential semantic drift that could occur for the images.

We recall that there is an original intention for each image, and that additional meaning accumulates with subsequent reuse/repurposing. An analysis of the situations that can arise within the archive shows that the original or current owner (the *custodian*) of an image can have one of four possible attitudes towards an image in their *hegemony*:

They could forbid reuse outright, for reasons of copyright or sensitivity.

They could forbid reuse of the image per se, but permit a copy to be made with a new life of its own.

They could be willing to share the image.

They could be willing to share the image and lose hegemony over it if someone else required it.

These attitudes would be expressed by a *usage status* indicator assigned by the custodian when the imaged was catalogued. There would also be a concomitant set of attitudes for the proposed reuse, depending of the intention of the *applicant* user.
This means that in any occasion of intended reuse, there would have to be a negotiation between the existing and intended usages of the image, leading to a protocol of reuse, using a hybrid system of versioning, messaging and automatic decision-making. We shall describe these next, using typical scenarios of use.

All reuse forbidden: Usage status FORBID

When reuse of an image is forbidden for social/legal reasons, there is no option. We call this image usage status FORBID (Figure 1).

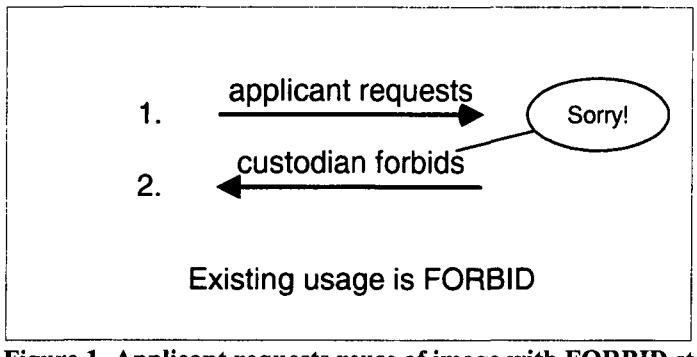

Figure 1. Applicant requests reuse of image with FORBID status; request is rejected.

This scenario has an automatic rejection of proposed new use: all applications are automatically rejected by the system regardless of their proposed reuse policy, with a pre-arranged message of explanation from the custodian to all potential applicants for reuse. As long as the custodian keeps up the FORBID usage status, this process is fully automated.
An example of this scenario would be where images have been collected by an ethnographer of culturally sensitive materials of a ritualistic nature; these are permitted to be used in the strict understanding that they won't be misinterpreted. Any reuse whatsoever, any re-situating through additional keywording or description will violate that condition. Therefore all reuse will be forbidden. Note that this does not apply to secret images or images that are not available to the public but to images that are available to the public but whose meaning cannot be altered: e.g. clustering of shamanistic spirits through aggregation of the totemic animals on zoological lines could be enormously offensive to the traditional owners.
On another level, family material in film and photograph could be given to a public archive, but only on the strict assurance that they were not used for public display, and never used in compilations.
In each of these examples, the custodian of the image is the archivist with responsibility for cataloguing the image in the museum, according to the wishes of the donor.

Reuse is only permitted of a copy of the artefact: Usage status RESTRICT

When an image serves an important role in an image database, and absolute control of that image is necessary, then reusing that image may be not desirable to its owner. However, if there is no other reason for this image not to be reused, then the image could be cloned, the original image keep its meaning, and entirely new image begin a new meaningful life. We call this image usage status RESTRICT.
In this case, there are two possible scenarios. In each, the applicant makes a request, and the custodian offers the possibility of a fork and the creation of a new artefact with new metadata to pass into the custodianship of the applicant. The two scenarios diverge with the decision of the applicant – they can either reject the offer (because they do not want to be custodians) (Figure 2) or they can accept it and repurpose the image with modified metadata (Figure 3).

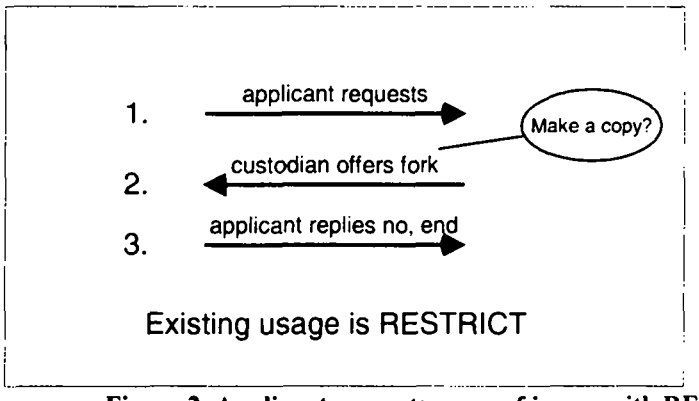

Figure 2. Applicant requests reuse of image with RESTRICT status and rejects offer of fork.

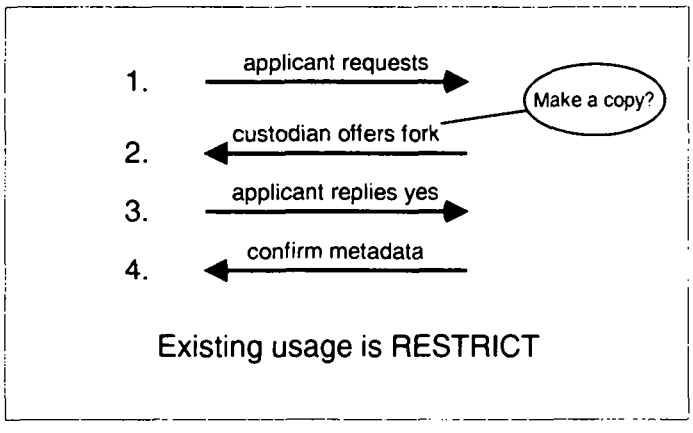

Figure 3. Applicant requests reuse of image with RESTRICT status and accepts offer of fork.

These two scenarios can no longer be automated completely, but only as far as the offer of forking: at this point the applicant has to make a decision based on his or her understanding of the role of custodianship of the media artefact, and then either accept or reject the automated offer.
An example of this could again be family photographs donated to the archive. A photograph of the family homestead could include factual details of the family and their history, but would be unlikely to include information of a symbolic or interpretive nature. Subsequent reuse and description of the image might lead to an unwanted associations: for example, the homestead in the photograph could be used to stand for a type of land use or farming style that was inimical to the actual owners of the homestead. Thus it would be desirable to ban reuse of the original image (so that, for instance, family details were not retrieved with the new usage), but repurposing with new metadata could be permitted.

Shared reuse may be negotiable: Usage status VOTE

The next usage status concerns the case where there is no need for restriction on the use of the image, but where the nature of any repurposing of the image can potentially lead to problems. Once the image is shared, the original custodian (by definition) no longer has total control over it, so that when any one of the users wishes to modify the image (crop, enhance, enlarge, shrink), modify its metadata, or completely replace it with another that can be substituted in terms of meaning, then that modification must be approved by all of the users for it to proceed. We call this usage status VOTE, since the mechanism of resolution employed involves voting on whether a change should happen or not. In this situation, the custodian serves as mediator.
When a modification/substitution is mooted (either by an applicant user, or else by one of the existing ones), there are two possible outcomes. Firstly, all existing users might consent – which results in the proposed change (semantic or physical) to the image being carried out (Figure 4).

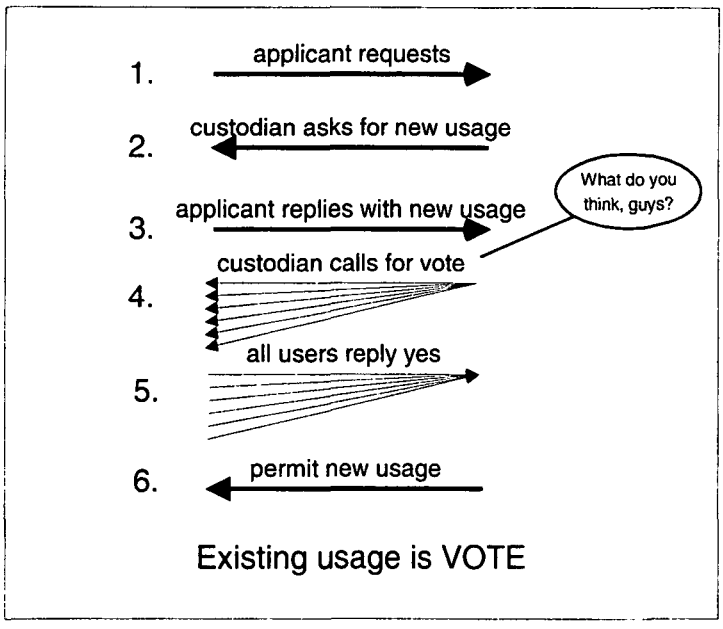

Figure 4. Applicant (new or existing user) requests reuse of image with VOTE status. All users reply Yes, so new usage is permitted.

Secondly, some or all of the existing users may dissent – which means that the existing image must be cloned (forked) and the copy modified. When this happens, the applicant/user may reject (Figure 5) or accept the forking, and in the latter case (Figure 6), some of the users of the old image may elect to go with the newer (modified) image, including the semantic implications that it carries.

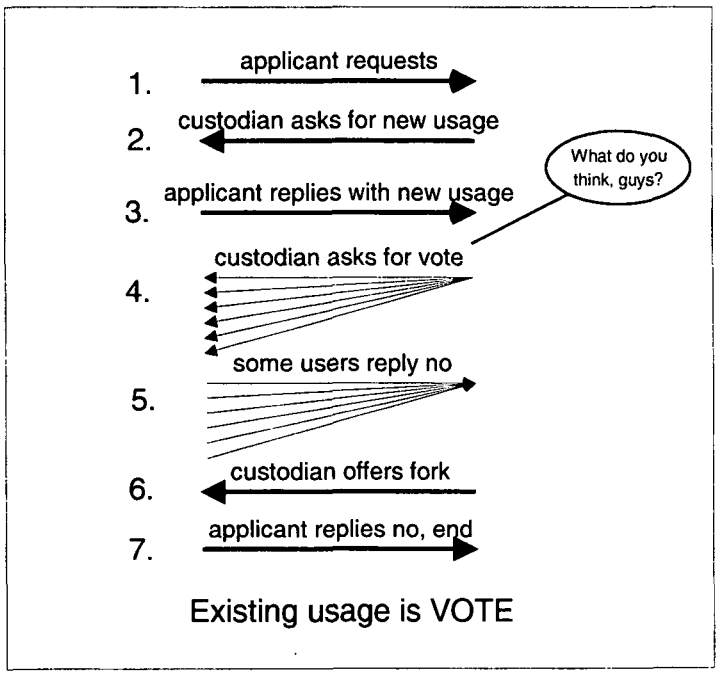

Figure 5. Applicant (new or existing user) requests reuse of image with VOTE status. Some users reply No, so custodian offers fork. Applicant rejects fork.

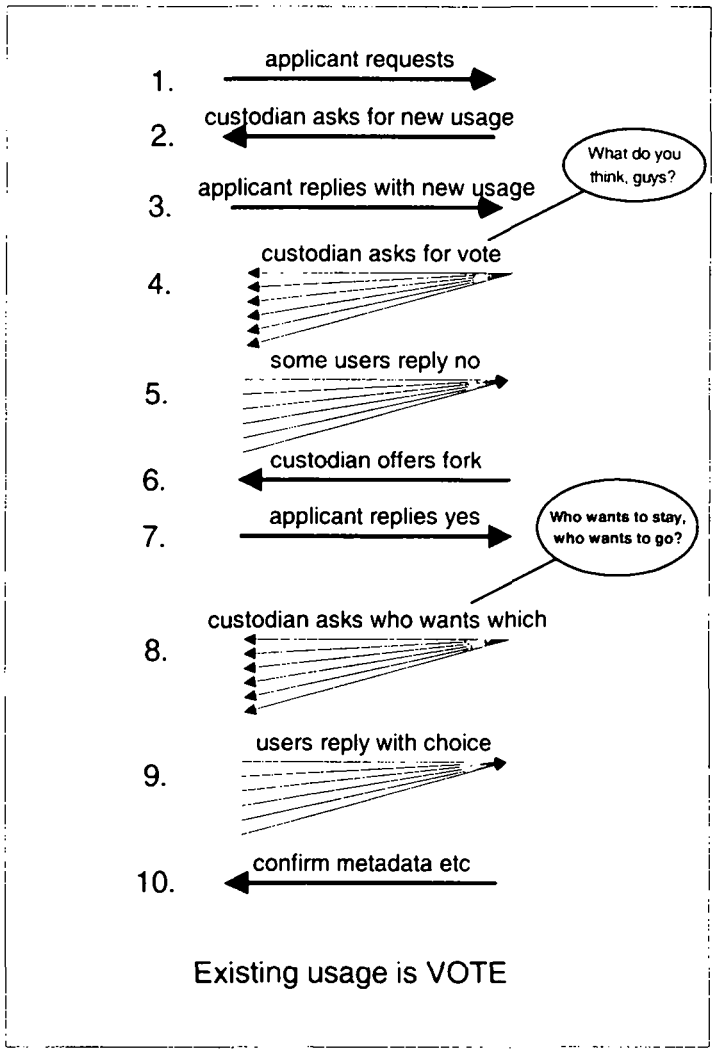

Figure 6. Applicant (new or existing user) requests reuse of image with VOTE status. Some users reply No, so fork is offered. Applicant accepts fork and users are polled to select which usage they wish to continue with.

An example of the replacement case might be where an image of a mountain was in use, and had the over-loaded semantic use of weather, tourism spot and geological formation. A better (or perhaps a clearer) image of "weather" may become available, and be mooted for replacement. However, the proposed replacement may not be good for illustrating one or more of the other existing uses. In that case, the forking would lead to a new image accession.
A more complex case of the VOTE usage status arises when an image has both representative and symbolic usages concurrently – the senses of *showing that* (an illustration of Mount Everest) and *standing for* (an illustration that mountains can arise from tectonics – Mount Everest rather than Kilimanjaro) lead to differing ways in which a new image might be substitutable for the image in current use. However, the protocol remains the same.
We note a special case of VOTE where the new usage proposed status is RESTRICT: here an automatic forking would occur, the processing of voting being unnecessary since the RESTRICT status denies future sharing.

The image can be shared, and custodianship transferred: Usage status SURRENDER.

When the custodian is the sole user of an image, but the usage is not one that can only be served by a particular image, the custodian may not only be willing to let the image be shared, but also to have another user take custodianship of it. This usage status is called SURRENDER (Figure 7).

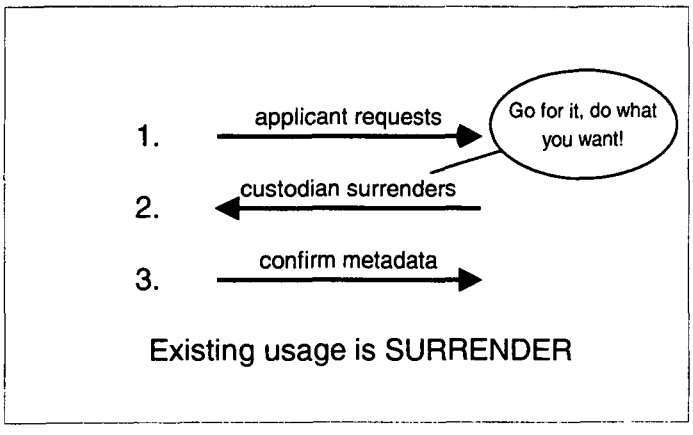

Figure 7. Applicant requests reuse of image with SURRENDER status and receives custodianship of image.

This case would most likely arise when the image in use is standing for a type, and the specifics of either image or metadata would not necessarily concern the custodian. A custodian who has an image of Kilimanjaro standing for "volcanic mountains" may be equally happy with a proposed replacement by an image of Mount Etna or Fujiyama. In effect, this case arises when a custodian does not want to take on the role of mediator in the community arising from subsequent image reuse. The original custodian remains part of the usage community for that image, wherein the new custodian mediates any VOTE decisions. A proposed replacement of the image by one of Mount Everest to illustrate "snow on mountain" would lead to the original custodian's requesting a fork, so that their usage context would not be violated.

DISCUSSION

The protocol proposed here would, when implemented in an enterprise media artefact repository, enable its semantically consistent reuse and repurposing. The hybrid social/machine rule-set described can effectively meet the situations that arise in the lifetime of such a repository, and could be implemented using a version control system in conjunction with a standard Internet RFC Call-for-votes system. We have described our proposed protocol using the simple example of an image archive, but the principles hold equally for digital assets of any media type.

There are several issues that remain the subject of further research. These issues include the potential trade-off between the lag-time in decision making and the possible urgency of a particular reuse, and the added complications that could arise from situations such as public archives where one part maintains use control, while another keeps copyright permissions. The preservation of media capital in an enterprise might require a DERELICT usage status to be assigned to artefacts that are not currently being used, so that their "usage pedigrees" are still available for subsequent mining of the metadata associated with them.

In particular, we intend to investigate specific communities of usage practices and their metadata requirements, including metadata to provide some indication of the reasoning behind the assigned usage status. It is anticipated that the in-community usage behaviour will lead to network effects arising from the interconnectivity of reuse links, and the community of knowledge they represent, while the existence of community standards may find that additional requirements on the reuse/repurposing (similar to those mentioned in the FORBID case) need to find expression in an additional set of rules.

REFERENCES

AAP. (Wednesday, June 6, 2001). Landmark case of Internet publishing. **Sydney Morning Herald.** Retrieved from: http://www.smh.com.au/news/0106/06/pageone/pageone6.html

Alavi, M., & Leidner, D. E. (2001). Knowledge management and knowledge management systems: conceptual foundations and research issues. **MIS Quarterly**, *25*(1), 107-136.

Bearman, D., & Sochats, K. (1996). **Metadata requirements for evidence**. Retrieved June 24 2001, from University of Pittsburgh, School of Information Science: http://www.lis.pitt.edu/~nhprc/BACartic.html.

Berners-Lee, T. (2000). **Weaving the Web: The Original Design and Ultimate Destiny of the World Wide Web**: HarperBusiness.

Bernstein, M. (2001). **HypertextNOW.** Retrieved from: http://www.eastgate.com/HypertextNow/archives/Cycles.html.

Besser, H. (2000). Digital longevity. In M. K. Sitts (Ed.), **Handbook for Digital Projects: A Management Tool for Preservation and Access** (pp. 155-166). Andover MA: Northeast Document Conservation Center.

Buneman, P. (1999). **Data provenance**. Retrieved from: http://db.cis.upenn.edu/Research/provenance.html.

Burnard, L. (1991). **What is SGML and how does it help?** TEI EDW25. Retrieved from: http://www.oasis-open.org/cover/edw25.html.

Chen, H. (2000). Digital libraries. **Journal of the American Society for Information Science Special Issue on Digital Libraries**, *51*(3).

Chen, S.-S. (1997). **Digital libraries: the lifecycle of information**. Columbia: Better Earth Publisher.

Clancey, W. J. (1997). **Situated Cognition: On Human Knowledge and Computer Representations (Learning in Doing)**: Cambridge University Press.

DCMI. (2001). **Dublin Core Metadata Initiative**. Retrieved from: http://www.dublincore.org/.

Digital Libraries Initiative. (2001). **Digital Libraries Initiative Phase 2.** Retrieved from: http://www.dli2.nsf.gov/.

Gammack, J. G. (1987). Different techniques and different aspects on declarative knowledge. In A. L. Kidd (Ed.), **Knowledge acquisition for expert systems: A practical handbook** . New York: Plenum.

Gammack, J. G. (1991). **Methodological issues in validating knowledge based systems**. Paper presented at the AMSE International Conference on Signals and Systems, Warsaw.

Gärdenfors, P. (2000). **Conceptual Spaces: The Geometry of Thought**. Cambridge MA: MIT Press.

Guha, R. V., Lenat, D. B., Pittman, K., Pratt, D., & Shepherd, M. (1990). Cyc: A Midterm Report. **Communications of the ACM,** *33*(8).

Habann, F. (2000). Management of core resources: The case of media enterprises. **The International Journal on Media Management**, *2*(1), 14-24.

International DOI Foundation. (2001). **The digital object identifier (DOI) system**. Retrieved from: http://www.doi.org/index.html.

ISO. (2000). **Overview of the MPEG-7 Standard** (ISO/IEC JTC1/SC29/WG11 N3445). Geneva: International Organisation for Standardisation. Retrieved from: http://www.darmstadt.gmd.de/mobile/MPEG7/Documents/W3445.htm

ISO/IEC. (1999). **ISO/IEC 13250 Topic Maps: Information Technology - Document Description and Processing Languages** (International Standard). Retrieved from: http://www.y12.doe.gov/sgml/sc34/document/0129.pdf

Lagoze, C. (2000, November 2000). **Business unusual: How "event-awareness" may breathe life into the catalog?** Paper presented at the Bicentennial Conference on Bibliographic Control for the New Millennium, Library of Congress.

Library of Congress. (1998). **Challenges to building an effective digital library**. Library of Congress National Digital Library Program. . Retrieved from: http://memory.loc.gov/ammem/dli2/html/cbedl.html.

Livemann, D. (2000). **Bacher caught in game of Chinese whispers**. CricInfo Archive, 21 April 2000. Retrieved from: http://www-aus3.cricket.org/link_to_database/ARCHIVE/CRICKET_NEWS/2000/APR/030523_CI_21APR2000.html.

Lynch, C. (1996, September 3, 1996). **Defining and maintaining attribute sets for use with the Z39.50 protocol: A discussion paper.** Retrieved from: http://lcweb.loc.gov/z3950/agency/attrarch/archive/cliff3.txt.

McCallum, S. (2000). **Extending MARC for Bibliographic Control in the Web Environment: Challenges and Alternatives**. Paper presented at the Bicentennial Conference on Bibliographic Control for the New Millennium, Library of Congress.

Nelson, T. H. (1990). On the Xanadu Project. **Byte** (September), 298-299.

Nelson, T. H. (1998). **Project Xanadu**. Retrieved from: http://www.sfc.keio.ac.jp/~ted/XU/XuPageKeio.html.

OCLC/RLG. (2001). **Preservation metadata for digital objects: A review of the state of the art** (White paper): OCLC/RLG Preservation Metadata Working Group. Retrieved from: http://www.oclc.org/digitalpreservation/

Polanyi, M. (1966). **The Tacit Dimension.** Garden City, N.Y.: Doubleday.

RDDL. (2001). **Resource Directory Description Language** *(RDDL)*. Retrieved from: http://www.openhealth.org/RDDL/.

Reich, V., & Rosenthal, D. S. H. (2000). **LOCKSS (Lots of Copies Keep Stuff Safe)**. Paper presented at the Preservation 2000: An International Conference on the Preservation and Long Term Accessibility of Digital Materials, York, England.

Sowa, J. F. (2000). **Knowledge Representation: Logical, Philosophical, and Computational Foundations**. Pacific Grove, CA.: Brooks Cole Publishing Co.

Star, S. L., & Ruhleder, K. (1996). The ecology of infrastructure: problems in the implementation of large-scale information systems. **Information Systems Research**, *7*(1), 111-134.

Svenonius, E. (2001). **The Intellectual Foundation of Information Organization**: MIT Press.

W3C. (2000). **Resource Description Framework (RDF) Schema Specification 1.0. W3C Candidate Recommendation 27 March 2000.** Retrieved from: http://www.w3.org/TR/rdf-schema/.