Putting the Gloss on Paper: A Framework for Cross-Media Annotation

Corsin Decurtins, Moira C. Norrie, and Beat Signer

Institute for Information Systems, ETH Zurich
CH-8092 Zurich, Switzerland
{decurtins,norrie,signer}@inf.ethz.ch

The final publication is available at http://dx.doi.org/10.1080/13614560410001725301

Abstract. We present a general framework for cross-media annotation that can be used to support the many different forms and uses of annotation. Specifically, we discuss the need for digital annotation of printed materials and describe how various technologies for digitally augmented paper can be used in support of work practices. The state of the art in terms of both commercial and research solutions are described in some detail, with an analysis of the extent to which they can support both the *writing* and *reading* activities associated with annotation. Our framework is based on an extension of the information server that was developed within the Paper^{++}project to support enhanced reading. It is capable of handling both formal and informal annotation across printed and digital medium, exploiting a range of technologies for information capture and display. A prototype demonstrator application for mammography is presented to illustrate both the functionality of the framework and the status of existing technologies.

1 Introduction

Despite predictions of the paperless office, paper and printed documents are still prevalent in current work practices [1]. For example, research scientists of all disciplines spend a lot of time *reading* and a lot of time *writing* and much of this work is based on paper. The reading activity may involve anything from the reading of scientific publications for review to the reading of images or statistical graphs for the purpose of interpretation. The writing activity ranges from filling in report forms to the authoring of reviews or publications. Annotations have been shown to be heavily used in the world of paper to enhance both reading and writing activities [2]. Further they often provide a basis for communication and collaboration whether it be as part of a co-authoring activity or publishing annotations to aid the interpretation of data or comment on its quality. There have been a number of proposals therefore to provide annotation tools in the digital world either as part of general web infrastructure [3] or for specific domains such as the annotation of brain images or tasks such as collaborative authoring or learning [4].

One particular form of digital annotation that has received a lot of attention in recent years is the use of annotations for semantic mark-up to assist in the interpretation of data either for purposes of automatic processing (e.g. in data integration architectures) or by humans. In particular, languages like RDF (Resource Description Framework) [5] and OIL (Ontology Interchange Language) [6] have been proposed for the expression of such forms of metadata annotations. Also under development is middleware such as SAM (Scientific Annotation Middleware) [7] to provide scientific researchers with a collaborative and cross-disciplinary working environment based around such annotations.

Our interest is to develop a general annotation framework that can support all of the many forms of annotation, not only within a digital world, but also across media. Specifically, we want to be able to support digital annotation of printed documents using technologies for digitally augmented paper. Over the past decade, there has been much research into such technologies and some are now available commercially, for example the Logitech io Personal Digital Pen [8]. However, it is important to point out that not all technologies have the same goal in mind and hence the same functionality.

We begin in Sect. 2 by discussing in more detail the various forms of annotation used and the role that paper can or does play to enhance the associated reading and writing activities. In Sect. 3, we discuss the state of the art in terms of both commercial and research solutions for digitally augmented paper, along with an analysis of the extent to which they support these activities. Section 4 then presents our annotation framework which is an extension of the information server developed in the European project Paper^{++}[9]. In the context of this project, we investigated the integration of digital and printed information with respect to enhanced reading. Support for the authoring activity is then discussed in Sect. 5. A prototype demonstrator for mammography that was used both to illustrate the functionality of the framework and also to evaluate existing technologies is presented in Sect. 6. Concluding remarks are given in Sect. 7.

2 Cross-Media Annotations

Annotation is some means of marking up a document so that it augments the existing material. For example, annotations could be used to explain terms occurring in a document, thereby providing direct access into a glossary during the reading process. Annotations come in many forms and have a variety of uses. They may be private or public, permanent or transient, and formal or informal. Informal annotations often take the form of free text but could also be sketches, images, or audio for voice commentary. For example, marginalia — the comments that we write in the margins when we read a paper — are informal annotations.

Formal annotations follow defined structures and conventions that enable them to be interpreted by other persons or computer programs. Typographic mark-up for the editing of documents is one such example, where it enables someone else to unambiguously interpret the changes to be done and carry them out. Another example of formal annotation that has been an active area of recent research within the scientific database and semantic web [10] communities is the use of annotation to mark-up digital data with metadata. This metadata describes the semantic content of the data and later enables either human readers or programs to access and process that data. For example, the annotation of scientific images with metadata can aid both search processes [11] and the integration of scientific data sources [12]. Further, within the hypermedia and semantic web communities, metadata annotation is also used to generate links between documents.

Most annotation tools support one particular form of annotation as they tend to support specific activities such as collaborative writing, data integration, search and so on. However, if one considers the working environment of most people, they have to work with lots of different forms of information and perform a wide variety of tasks that may require different forms of annotation. For example, in a study of university textbooks, it was shown that some kinds of annotations were used to support *reading*, while others were used to support *writing* [13]. In the case of reading, notes are often written alongside text and figures to aid interpretation in future readings (either by the same reader or other readers) and important sections highlighted or underlined.

Another study examined the task of writing document summaries and how users would annotate the document and take notes [14]. In this case, readers not only wished to highlight important items, but also to extract and re-order them according to the final structure of their summary. Annotations alongside the text heavily used references to structure outlines. Further, this study examined the differences between performing this task using only digital documents as opposed to the use of only paper. It was shown that there are many problems with digital annotation systems in terms of both inputting the actual annotations and also working with various documents alongside each other.

Within the scientific domain, one frequently finds that paper forms are heavily used as a means of both collecting data and also reporting on the analysis of data. For example, in breast screening clinics, mammograms are analysed by experts referred to as "readers" to determine whether or not the patient should be called for further tests. For each patient, four mammograms are taken — two views of each breast — and the resulting film is analysed by a reader and the results reported on a paper form. The form contains breast outlines corresponding to the four mammogram views and the reader will annotate these to indicate any abnormalities or special features and, generally, any reasons for their decision whether or not to recall the patient.

It is common in most clinics to have *double reading* to reduce the chances of missing a possible tumour. In many cases, the second reader will first analyse the mammograms and annotate the form before then checking the analysis of the first reader. This then gives them the chance to double check whether or not they have overlooked something. The ability to read the annotations of other readers also plays an important role in helping less experienced readers learn from more experienced ones. Further details of the practices of breast screening clinics and their use of annotations are given in [15].

One could propose that the whole breast screening process be digitised. Mammograms could be scanned and digital versions generated. Instead of using paper forms, the readers could enter data in electronic forms and annotate the forms and/or the digital versions of the mammogram digitally using some kind of electronic notebook. However, this scenario presents a number of problems. First, the clinics currently see the generation of digital versions of the mammograms as simply creating an extra step in the process, resulting in extra delays and costs. Second, it is much more comfortable and quicker for a reader to enter data on a paper form, especially the annotations. It is also interesting to see that readers will often work in batches, analysing a number of mammogram sets until a significant feature is detected and then filling in the forms for the whole batch.

This is just one example of many that we have encountered, where users are very reluctant to move away from paper. Sometimes it may be part of general resistance to change, but studies of the working environment and work practices can often reveal that there are very sensible reasons to remain with paper and other forms of non-digital media. On the other hand, if we consider the breast screening example, there are certain activities such as the reviewing of previous reports and annotations that could be improved with digital support. In particular, if a second reader or a trainee could selectively reveal annotations, it could be extremely beneficial. Last but not least, it is usually desirable to use digital media for the long-term storage and management of data which means that there needs to be an easy and flexible way to digitally capture the information on the paper form.

Paper as a medium has many advantages over digital media in terms of how people can work with it, both individually and in groups. It is portable, cheap and robust. It is much more convenient to scan through a book by rapidly flicking through pages than to browse a digital document. Paper also supports forms of collaboration and interaction that are difficult to mimic in current digital worlds. Whereas the focus in the past has been to replace paper, increasingly, the trend is towards integrating printed and digital media, thereby achieving the best of both worlds. The rest of this section will introduce two different forms of annotating paper documents and outline the requirements for augmented paper.

The first form of augmentation we want to support is *enhanced reading*, whereby users are supported in their reading activity. This means that they should easily be able to navigate within an information space that spans not only different information sources, but also different media. Specifically, we want the readers of printed documents to have easy access to digital annotations. The reader should be able to simply point to an active area within a page and immediately view the linked information, possibly following links to further digital materials and even back to paper. This scenario requires a reader device that has the basic capability of sensing and immediately transmitting position information. The emphasis here is on *interaction* rather than *capture* and the desire to turn paper into an interactive device.

As an example, we present various forms of annotation of printed materials in Fig. 1. On the left, we show different parts of a printed table annotated with digital content. The bar chart refers to the whole table, whereas the pie chart focusses on a specific column. Last but not least, there is a text note on a specific table entry. This ability to annotate different granularity levels is also an important feature. Further, it is important to support both personalisation and sharing of annotations in a way that information delivery is adapted according to both the user and the context. On the right of Fig. 1, we show various forms of annotation of parts of an image, including a video, a text and a web page.

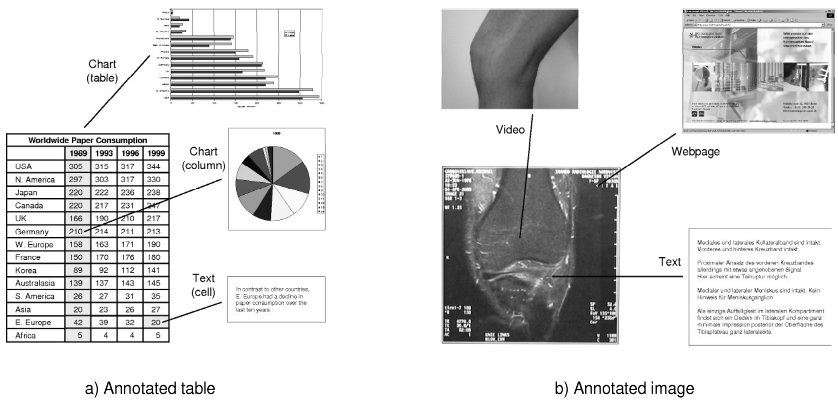

Fig. 1. Annotations of printed materials

The second form of paper augmentation involves an active capturing of information and enhances the process of writing. The *enhanced writing* can be regarded simply as linking pieces of information, where the link source specifies what is being annotated and the link target is the annotation. Annotating differs from normal linking through the fact that the authoring of the target resources is often an important part of the annotation process. Thus, in addition to *link authoring*, annotating also includes *content authoring*.

For enhanced reading as well as enhanced writing we need a way to bind digital or physical entities to parts of a paper document. Further, a user should be able to activate (trigger) a link and access the information stored as an annotation. Enhanced writing further requires a way of acquiring new information. The information capturing process may be based on regular keyboard input or form an integral part of new writing tools (e.g. digitally enhanced pens).

In the next section, we introduce various technical solutions for integrating printed and digital material, including our own solution developed as part of the European project Paper^{++} [9] which focussed on enhanced reading.

3 Digitally Augmented Paper

Over the last decade, various solutions for augmented paper have been realised and successfully applied in demonstrator applications. In this section, we start by giving an overview of past projects and then go on to discuss state of the art technology and its application in commercial products. Note that since we are interested in addressing parts of documents and not only documents as a whole, we will not discuss technologies used for document tracking such as Radio Frequency Identification (RFID) tags [16], CyberCode tags (visual 2D-barcodes) [17] or Infrared Beacons [18].

The Digital Desk built at Xerox EuroPARC brought additional functionality to an ordinary physical desk [19, 20]. A camera mounted above the desk surface is used to track a user's interaction with documents lying on the desk. Additional information can either be displayed on a separate computer screen or, in a more advanced version of the digital desk, a projector is used to directly display information on the table and overlay the physical paper. Since the paper documents always have to be processed on the desk to get access to the supplementary functionality, we lose one of the main benefits that paper affords in terms of its mobility. However, for fixed installations such as museum exhibitions, the digital desk metaphor still has a lot of potential. Recent work on augmented desk interface systems focusses on real-time hand tracking and gesture recognition [21] and the development of applications enhancing working activities such as architectural design [22].

Instead of having a fixed installation, we can bring more functionality to the tools (e.g. pens) used when working with paper documents. A highlighter pen combined with a small camera (VideoPen) has been used in the PaperLink system [23] to augment paper documents with electronic features. PaperLink uses computer vision and pattern recognition techniques to detect specific printed words on a page and to link them to some digital content. However, PaperLink's main focus is on supporting the definition of links from paper to existing digital content rather than pen-based capturing of written information. The related digital content has to be chosen in an application specific way (e.g. by selecting a file in a file chooser). Therefore, we would classify PaperLink as a representative of an enhanced reading system.

PaperLink uses existing information printed on paper as anchors for links to digital information and this is just one possible solution for integrating printed and digital material. DataGlyphs [24, 25] developed by Xerox can be used to store additional digital information on paper using a special printed pattern. In this case, forward and backward slashes representing zeros or ones are used as a digital encoding. If the pattern is small enough, it is not intrusive to our eyes. The encoded information can be extracted by a scanning device and could be used as a means of invoking digital services.

On the other hand, there are hardware solutions which are dedicated to information capturing such as the mimio Xi [26]. The main idea is to enhance whiteboards with capturing facilities. A pen's position is tracked by a high-resolution ultrasonic position detection device and handwritten information becomes automatically digitised. An interesting feature of the mimio Xi solution is that the capturing process is independent of the actual medium on which the pen is used. Therefore, the mimio Xi technology is a candidate to augment media other than paper, such as X-ray films used in mammography. The InkLink handwriting system from Seiko is a smaller and more portable version of the mimio Xi, which can be used for documents up to A4 paper size. The mimio Xi and InkLink are well suited to information capturing. However, there are some problems in using them for enhanced reading because a document always has to be calibrated with the reading device before supplementary digital information can be accessed. Also, these technologies are essentially page-based rather than document-based. They do not provide any mechanism that allows multiple pages to be associated with a single document.

A technology which can potentially support enhanced reading as well as enhanced writing has been developed by the Swedish company Anoto [27]. Again, the idea is to get a pen's x and y position on a paper document, but without a preceding calibration procedure. Therefore, the position information is directly encoded on each piece of paper, in this case using a special pattern of tiny visual dots. One can assume that we have a virtual grid over a page and the dots are placed relative to the intersections of the horizontal and vertical lines. Each dot then encodes a two bit sequence which is defined by its displacement (horizontal and vertical offset) from the corresponding intersection point. Several dots together form a unique sequence of zeros and ones which finally defines a position in a large virtual document space. The dot pattern results in a slightly grey page background with minimal interference with the document's printed content. A special reader pen for the Anoto dot pattern has to be equipped with a camera next to the writing stylus to track the pen's movement relative to the paper surface. A record of the pen's movement can then be used to recreate what a user has written within the digital world. Capture of additional information such as the pressure and angle of the pen can be used to enhance the rendering process and create images that really recreate the look and feel of hand-written text and sketches.

Several Anoto pens from Sony Ericsson (Chatpen), Logitech (io Personal Digital Pen) and more recently from Nokia (Digital Pen) are now available on the market. Currently, there is a clear focus on information capturing since a user's actions (strokes) are stored within the pen and only transmitted to another device on demand. However, this technology has the potential to be used for enhanced reading and direct interaction: It requires only that the position information be transmitted continuously rather than in batch mode.

In addition to the three pens just mentioned, there are other Anoto pens which provide different functionality. For example, the MyPen by C-Channel can read the Anoto pattern (position information) but also works as a capturing device for printed text (e.g. to read the account numbers on pay-in slips). It is tethered to a computer and immediately transmits new pen input. The pen is sold together with a mouse mat with predefined "active areas". If the user points to an active area covered with the Anoto pattern, a specific action will be triggered. In the case of the mouse mat deployed with the MyPen, users can define physical bookmarks for favourite websites which will be loaded as soon as they point to the corresponding active area. Further, there is a larger area in the centre of the mouse mat that can be used to move the mouse cursor on the computer screen. If we combine the MyPen's facility for direct transmission of positional information with the capturing facilities of the Anoto pens presented before, we would get a pen supporting enhanced reading as well as enhanced writing.

Note that, not only is the Anoto technology applied in current research projects [28], but also commercial enterprise software solutions have been realised using it as the driving hardware technology. For example, Hewlett-Packard is selling a forms automation system which integrates data filled in a form into a company's business database.

Finally, we would like to introduce a solution for augmented paper developed within the European project Paper^{++}in which we participated. The goal of the project was to investigate concepts and technologies for enhanced reading in everyday settings. Cost was a major consideration in the technology development as we sought solutions that could be deployed widely in schools, homes etc. Therefore the cost of reading devices should be so low that they would almost be at the level of disposable technologies, i.e. only a few Euros. Consequently, it was necessary to devise a solution that avoided expensive optical components such as the camera used in Anoto pens.

Our project partners have developed a position encoding pattern and built the corresponding reader hardware. In contrast to the visual encodings of Anoto's solution or DataGlyphs, the Paper^{++}encoding is based on a grid of invisible barcodes printed with conductive ink. A specially designed pen reads the information encoded in the barcodes by measuring the inductivity and this information is decoded to get the corresponding x and y positions. The current Paper^{++}solution therefore is effective for enhanced reading but does not support pen-based information capturing.

4 Cross-Media Annotation Framework

In the previous section, we described various technologies that can be used to achieve integration of printed documents and digital information. Currently, this is a very active area of research and surely more technologies will become available in the near future. However, we believe that the key to developing powerful cross-media annotation systems lies primarily with the information server rather than the specific client technologies. Further, it is important that a general cross-media annotation framework be independent of the particular technologies for digitally augmented paper. In addition, it must be flexible and extensible with respect to media types in order that it can support the various forms of annotations discussed in Sect. 2, as well as possibly new unanticipated ones that might emerge in the future.

We now present our cross-media annotation framework. In contrast to a number of other annotation architectures e.g. SAM [7] and COHSE [29], our general approach is based on extending database technologies through the integration of new concepts, rather than on a middleware approach based around existing database technologies. Also, many systems are available that support either formal or informal annotations, but there are few systems that support both forms, mixed forms and allow navigation back and forth between structured and unstructured annotations. We believe that this is one of the crucial issues for a generic annotation system. Another important feature is the integration of various media types, both digital and physical.

The annotation framework is embedded in the overall system architecture shown in Fig. 2. To abstract from the particular devices and technologies used for digitally augmented paper, a special Input Device API has been developed as part of the Paper^{++}project. New devices can easily be integrated by using this API and can immediately make use of the framework. This means that, in general, we do not depend on a specific device for the authoring of annotations on paper documents and, as indicated in the figure, we currently support access based on Anoto, mimio/Seiko and Paper^{++}technologies.

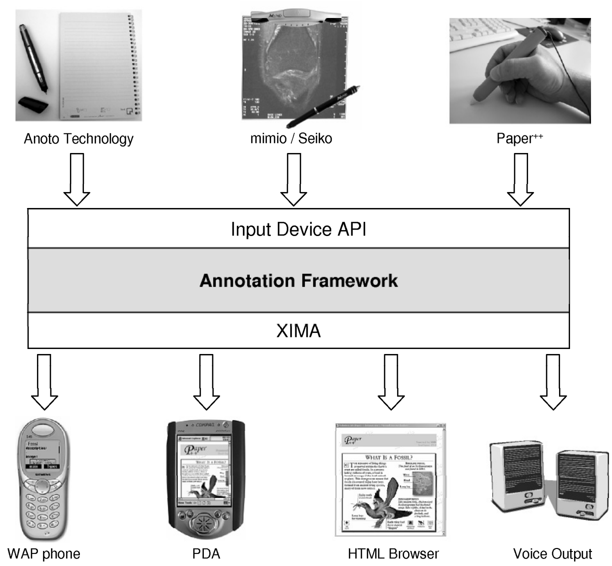

Fig. 2. Architecture overview

Display devices access the annotation framework through a flexible content delivery platform called XIMA [30]. It integrates various output devices including web browsers, desktop computers, cell phones, PDAs and speech output. The availability of XIMA means that it is relatively easy to, not only integrate visualisations for new media types into the annotation framework, but also to introduce new client display technologies.

Figure 3 provides an overview of our annotation framework. It is based on the information server that we developed in the Paper^{++}project. This server, known as iServer, is an open hypermedia system that supports various media types, including paper documents. The core of iServer is a generic link framework that is independent of the specific media, with the various resource types being integrated through a plug-in mechanism. In the current implementation, it supports paper documents and databases as resource types in addition to uninterpreted digital resources such as images, videos and web pages.

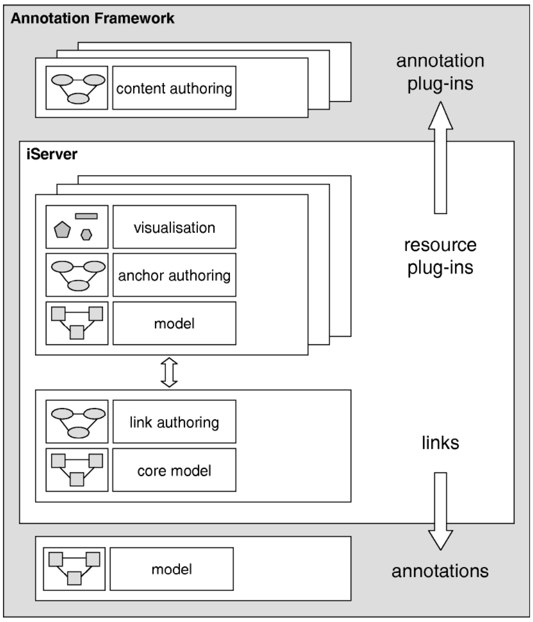

Fig. 3. Annotation framework

Here we use the term *uninterpreted resource* to denote a resource that can be referenced by the framework, but is interpreted only by external applications used to view the resource. In contrast, an *interpreted resource* is one for which the framework understands the structure and is able to reference elements within the resource. At the moment, both paper documents and databases are interpreted resources as the framework can reference elements within a paper document through the notion of active areas and also objects within a database through a querying mechanism.

The reason for using iServer is that annotating an artefact basically corresponds to the authoring of a link in an open hypermedia system along with the authoring of the content to which the link points. However, note that, instead of adding new content, users could also make annotations by linking to existing resources. Therefore, the difference between annotations and links is actually quite subtle. Besides the fact that the process of annotating frequently includes adding new content to the system, we regard an annotation as basically being just a special classification of a link. Classifying them as annotations simply provides an easy means of handling annotations in a special way, such as making them visible or invisible depending on whether we want to see the augmented document or just the original. By using iServer as an underlying system, a lot of the requirements are already met. iServer provides APIs and a data model for both generic links and extensions for specific resource types as already discussed. It is out of the scope of this paper to give a detailed explanation of the iServer architecture. However, we would like to outline some of the iServer's main features.

In Fig. 4, we present the main components of the iServer architecture. `Links` are first class objects which can have one or more sources and lead to one or more link targets (*multi-headed links*). By modelling them as a subclass of `Entities`, we achieve full generality of allowing links over any type of entity objects, including links themselves. Further, we do not distinguish between components which can be used as link sources and those applicable as link targets. Every object that can be used as a link source is also a valid target component and vice versa. The next type of entity is the `Resource` type representing an entire information unit. By introducing the resource concept we can, for example, link from an entire HTML document to a single movie clip. However, quite often we would like to control the linking granularity by being able to address a specific part of the document rather than the document in its entirety. Therefore, as a last subtype of the entity type we offer the concept of a `Selector`, a construct enabling parts of the related resource to be addressed (similar to the reference objects described in the FOHM model [31]).

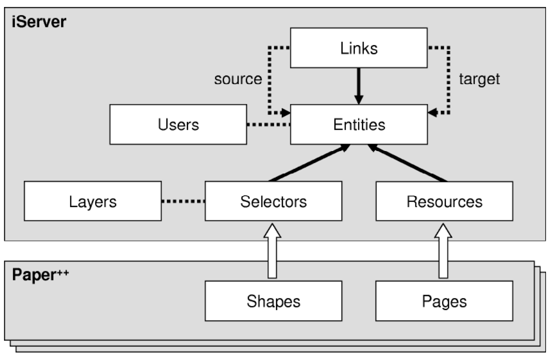

Fig. 4. iServer with Paper^{++} plug-in

The introduction of the selector concept provides a mechanism to control the granularity of link sources and link targets. But did we really achieve the desired flexibility in addressing arbitrary parts of a resource? The answer is *no* since there is still an important concept missing. What happens, for example, if we define one selector addressing parts of a resource and a second selector addressing a subpart of the first selector's range? For example, one selector might address an entire text paragraph, while a second selector addresses a phrase within that paragraph. Then if an input lies within the range of both selectors, for example, a word within the phrase and hence also the paragraph that contains it, it is no longer defined which link should be activated. To handle overlapping and nested link sources or targets, the iServer architecture therefore provides the concept of `Layers`.

A resource may have any number of virtual layers. Each selector is associated with exactly one layer of a resource and the selectors on any given layer must be non-overlapping. Then, for a given input, we detect the selectors for which that input belongs to the selector's range. If more than one selector is returned, then they must belong to different layers and, depending on application semantics, the appropriate layer and hence selector is activated. Layers are ordered and may be re-ordered dynamically by the application, which may also activate and de-activate individual layers at any time. Therefore, while the general rule is to activate the selector on the "uppermost" layer, there is great flexibility in how the application can control this depending on various factors such as the history of activation and user task. Further details of the layer concept together with the link model in general are given in [32]. Other hypermedia systems such as Hyper-G's Harmony browser [33] support overlapping link anchors but they do not provide any functionality to explicitly define the semantics of such overlapping links.

In addition, to define the link behaviour, we can associate every link with a set of properties. These properties consist of key/value pairs which are application-specific rather than predefined by the iServer framework and therefore can be applied to customise the link behaviour for specific application domains. To give an example: we could define a property named *onActivate* which would define the action to be taken when a link has been triggered. Possible values could be *openInline* to open the link target within the current component or *openNew* to present the target separately from the link anchor. Similar concepts exist for example in the XML Linking language (XLink) [34] where the *actuate* attribute is used to define the traversal behaviour and the *show* attribute defines where a link should be shown (e.g. replace, new window). However, we try to be as flexible as possible by not predefining a set of properties but rather introducing an abstract property set which can then be extended by specific applications.

Various open hypermedia frameworks, for example the Devise Hypermedia system (DHM) [35], an extension of the Dexter hypertext reference model [36], have been proposed. The sharing of link knowledge within these frameworks, and collaboration between different distributed link services and users, has always been a goal of the open hypermedia community. However, most open hypermedia models and architectures, including more recent ones such as the Fundamental Open Hypertext Model (FOHM) [31], do not consider user management and the issues of *data and link ownership* as core services. In contrast, we regard link ownership as critical, especially in open and distributed environments, and therefore provide user management as a core component of our architecture. Every entity (resource, selector or link) has an explicit owner (creator) who can define the entity's visibility to other users. Note that by positioning the user management at the level of entities, we get the flexibility to handle access rights not only at link level but also on the base of a link's source or target objects. An application of the user management in collaborative annotation will be described in Sect. 6.

The iServer link metamodel has been directly implemented on OMS Java, a persistent, object-oriented data management framework [37]. OMS Java is an implementation of the OM model [38] which differs from commonly used object models such as UML in that it is, not only intended for system design, but also as an operational model for data management. Thus the OM model defines a full operational model over objects, collections (both ordered and unordered) and associations as well as constructs for their definition. Instead of defining a complex hypertext architecture and then storing all the link information in a separate database, or even a file system, we directly empowered our database objects with the required hypertext functionality. Currently, iServer provides a Java API and an XML interface which can be accessed locally or contacted remotely by client applications using the iServer Java Servlet gateway.

The lower part of Fig. 4 shows a specific resource plug-in which has been developed as part of the Paper^{++}project to integrate paper and digital information (digitally augmented paper). Link source anchors within a printed document page correspond to *active areas* and therefore the specific selector for digitally augmented paper is represented by `Shapes`.

Table 1 provides an overview of current iServer plug-ins together with the corresponding resource and selector types. For example, the existing movie plug-in allows the use of time spans as selectors for movie clips. However, another application might need to link movies based on spatial information within the movie and a third one based on a combination of temporal and spatial information. The iServer architecture therefore allows a user to define different selectors for the same media type.

Medium	**Resource**	**Selector**
webpage	XHTML document	XPointer
movie clip	mpeg file, avi file etc.	time span
sound	mp3 file, wav file etc.	time span
augmented paper	document page	shape
database	database workspace	database query

Table 1. Plug-ins

The resource-specific plug-ins of iServer include extensions of the generic data model's concepts for resources and selectors and application components for the visualisation of the resources along with the source anchors and links. The visualisation component for the image plug-in, for example, displays the image document overlaid with semi-transparent shapes representing the source and target anchors of the links. These links can be activated by clicking with the mouse on the active areas. Note that, in the case of the Paper^{++}plug-in, the visualisation component is the digitally augmented printed document itself. Anchors can be visualised in various ways (visible or invisible) and can be activated by the appropriate reader devices. As shown above, so far we have developed plug-ins for XHTML, audio, video and still images that will enable elements within web pages, sound clips, videos and images, respectively, to be linked in addition to entire resources.

The authoring of the links can be covered in a generic way by the iServer core system, but the anchoring of the links in a resource is implemented by the specific resource type plug-ins. Some of the plug-ins also come with a component for the authoring of link anchors. In the Paper^{++}plug-in, for example, this is an application that uses PDF versions of the printed documents. It displays the documents on a screen and the user can draw the anchors on top of the documents with a mouse. For the annotation framework, we have developed an additional anchor authoring component for paper documents, that lets users select the anchor on the paper document itself.

In addition to the visualisation and link authoring components defined by iServer, other components for the authoring of content are added in the annotation framework. These components are obviously also specific to the resource type. In particular, we have focussed on a component for the authoring of text annotations and sketches on paper. More details about authoring of links and content on paper documents are given in Sect. 5.

Figure 5 gives an overview of our annotation model. In the centre we show the main components of the iServer model. It contains the generic modules `Links` and `Resources`, which can be extended for specific resource types through the `Plug-ins` module. Note that the iServer model also contains a `Users` module for information about users, user roles and access rights. Further, personalisation and authorisation of information access is based on this module, something that is often not supported in annotation systems.

Earlier in this section, we discussed the relationship between the concepts of links and annotations. In our framework, we have modelled `Annotations` as a subcategory of `Links`, i.e. links can be classified as annotations. This design leaves a maximum amount of flexibility to the application and allows for a tight integration of link server, open hypermedia and annotation issues. The classification of annotations can be refined even further by introducing subcategories (e.g. `Comments`, `Explanations`, `Examples`, `Formal` or `Informal`). Note that an annotation can be classified in multiple categories at the same time and applications can define their own categories.

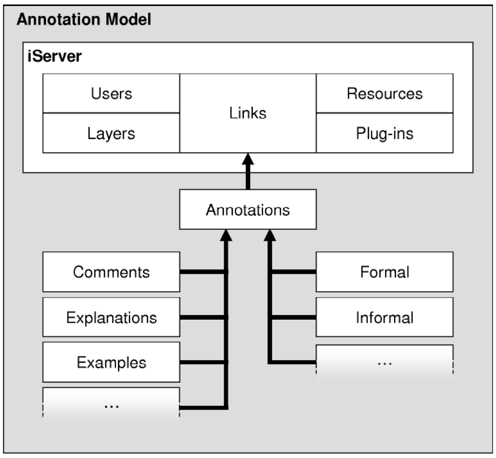

Fig. 5. Annotation model

Having presented our generic cross-media annotation framework, in the next section we go on to address issues of the authoring of annotations, especially through the medium of printed documents.

5 Authoring

When analysing the activity of authoring annotations, there are two main factors to consider. First, there is the issue of *annotation capture* in terms of how users author the actual content of the annotation. Second, there is the question of the means used to define an element within a resource and link it to an annotation or concept.

iServer and its media plug-ins already provide APIs for adding new resources such as videos, databases and printed documents, and defining anchors on these resources in terms of selected elements within the resource. For some media types complete authoring applications are available. These components can also be used for the authoring of annotations and anchors for the links. For example, for paper documents, an authoring application has been developed that allows the authoring of links for pre-printed paper documents. The documents are available electronically as PDF files. Anchors on these documents can be defined through a graphical user interface and can be linked to existing resources, for example, to objects in a database.

However, in the case of paper documents, it would be useful to author links and annotations directly on the paper document. As described in Sect. 3, various devices for capturing writing from digitally augmented paper have recently become commercially available and it is possible to use these for both content and link authoring on paper and similar physical media.

Anchors for the annotations can be defined by marking a specific area on the paper document with the writing device. *Unstructured* annotations such as handwritten text or sketches can also be captured and digitised by the pen. The captured content is transmitted and added to the annotation server. Some of the current devices provide simple character recognition for handwritten text, but these systems are not very advanced and sophisticated yet. Another option would be to run character recognition software on the PC to which the pen is attached. However, in practice there are many applications where the annotation does not need to be interpreted and it may therefore be sufficient to store the handwritten text as an image.

The pens can also be used for *structured* annotations, e.g. for filling in forms. The Logitech io Personal Digital Pen for instance is even able to store information about form fields in the pen itself. Unfortunately, this information cannot be modified by the user. However, the form semantics can also be implemented in the server. Then instead of just storing the captured content, it is analysed and matched against the definition of a form. Checkboxes, sliders, text fields and similar components can easily be implemented. For some of the fields, it might even be feasible to apply handwriting recognition. The limited amount of content and prior knowledge of the content type, e.g. that the content will be a number or a date, can significantly improve the accuracy and performance of the handwriting recognition. The information extracted from the forms can then be used to create structured objects as the content of the annotation.

If the pen is used to author both anchors and content at the same time, the system has to be able to distinguish between these modes. There are several possibilities as to how this can be done. One is that the anchors and annotations are made on two separate pages. This is the configuration that we have used for our mammography prototype. The anchors are made on the X-ray film while the annotations are made on a separate report form. Which anchor belongs to which annotation is determined by the sequence of the entries. This makes it quite easy for the application, but the users have to be very disciplined in the authoring process. We have investigated other input modes to enable us to compare approaches. For example, we also implemented a version with the anchor definitions and annotations on the same page, using predefined checkboxes to switch between anchor definition and annotation mode. At the moment, we have carried out only preliminary studies and it is too early to say which method is preferred. Also, a variety of other input modes with non-sequential authoring of anchors and annotations are possible and should be investigated further. However, it is worth noting that all of these solutions tend to have similar advantages and disadvantages.

In authoring activities, it is necessary to give feedback to the user. For example, if the content of a form field could not be captured correctly, a user should be notified about the problem. In the case of combined anchor and content authoring, the user should know whether the pen is currently in the state for anchor authoring or content authoring.

Most of the existing Anoto pens can give feedback through vibrations. Some also have LEDs, but these are less useful as users tend not to look at the pen when writing or sketching. The problem is that most pens are not interactive, i.e. they send data to the computer when they are placed in a cradle or a special button is pushed. This means that information capture tends to consist of two phases — a writing and a transfer phase — with all data being sent as a single unit. This is similar to early web-based form processing where all input data was validated and processed in a single step, rather than being able to validate and act upon individual data values as data is entered. For immediate feedback in the case of the Anoto pens, the annotation component would have to be able to get the information captured by the pen immediately and send back feedback. With the pens available currently, this is not possible. However, it is important to point out that existing pens have the potential for this: It only requires that they be adapted so that they could also be used in an immediate data transfer mode. Based on our discussions with the technology developers, we anticipate that this option will be available in near future.

To investigate the approach of annotation and link authoring on paper, as well as the needs and possible benefits of cross-media annotation in scientific environments in general, we have developed a prototype system to support the work of the mammography screening process mentioned in earlier sections of this paper. It uses the Logitech io Personal Digital Pen for the authoring of both links and annotation content and described in the next section.

6 Mammography Application

In breast screening programmes, women are checked for symptoms of breast cancer on a regular basis. Multiple X-ray images are taken of each breast. The images are then analysed by doctors and, if suspicious features are found on the images, the patients are recalled for further testing. The process of "reading" the mammograms i.e. analysing the X-ray images, has been investigated and described in [15].

This is a typical example for the use of annotations in the context of scientific data. Paper and other non-digital media play a very important role in this process. The X-ray images are available on film and the readers use a paper form to report on the results. Some breast screening centres have already switched to digital technologies or are considering it. The X-ray images can be displayed on large computer screens and computers can be used to report on the X-ray images. The big advantage of this purely digital approach is that the reporting data is available in a digital form for querying, ubiquitous access and for archiving. Researchers are also working on automated analysis of the images. However, switching to digital media results in a significant change of the working process and it appears that it also brings disadvantages [39]. We believe that an integration of paper and computer systems is much better than simply replacing paper by digital media.

In our mammography prototype application, we use the technologies described in the previous sections to improve the process of mammography reading. The aim is to interfere as little as possible with the current work practices, but to provide users with additional functionality.

The same technologies used for augmenting paper documents can also be applied to other non-digital media. For the mammograms, the mimio Xi product [26] described in Sect. 3 could be used for analysing banks of X-ray images and the Seiko InkLink for analysing single images. Both technologies detect the position of the pen with high-resolution ultrasound without requiring changes to the X-ray film. Since we only have access to limited X-ray images on film, we also used paper copies as substitutes in our prototype, using the Logitech pen for both the mammograms and the report form. The pen can store input from multiple pages simultaneously and the data is transferred to the computer when the pen is placed into its cradle. The data is stored in the form of an XML document for each page. The XML document contains all the strokes as a sequence of points (x and y positions) along with additional data and metadata. Since we currently do not have access to the low-level API, we work with these XML documents. The annotation prototype is based on an OMS Java database [37] that includes an application database with information about patients, screenings, screening reports, doctors etc. as well as the annotations.

The mammography reading workplace is shown in Fig. 6. For each patient, the reader has multiple sheets of paper, covered with the Anoto pattern — the mammograms of the patient and a screening report form. The report form contains an active area linking to a patient's medical record. As discussed in previous sections, the current version of the Logitech pen and other similar products do not support interactive working and the pen has to be placed in the cradle after every selection.

To fill in the report form, the reader marks checkboxes and writes free text in the corresponding areas. After placing the pen in the cradle, the content of the checkboxes is analysed and attributes of the corresponding objects in the database are set. The free-text annotations are stored as images and added to the database as an annotation of the mammography. Note that we do not apply any handwriting recognition since it is a very expensive and error-prone process. The free text annotations will only be interpreted by the other human readers of the mammogram and so there is no need to store the annotations in a machine-readable form. Any information that has to be interpreted by a program is entered using a checkbox form field.

If a reader wants to annotate a specific area of a mammogram, he can use the pen to draw a shape around it before entering the annotation text on the report form. The system creates an active area on the mammogram in the form of a polygon and links this area to the digitised annotation text, thus creating a digital annotation to the paper document. This gives the readers the possibility of creating annotations on different levels of granularity. They can annotate the whole mammogram as well as very specific details referring to parts of the images.

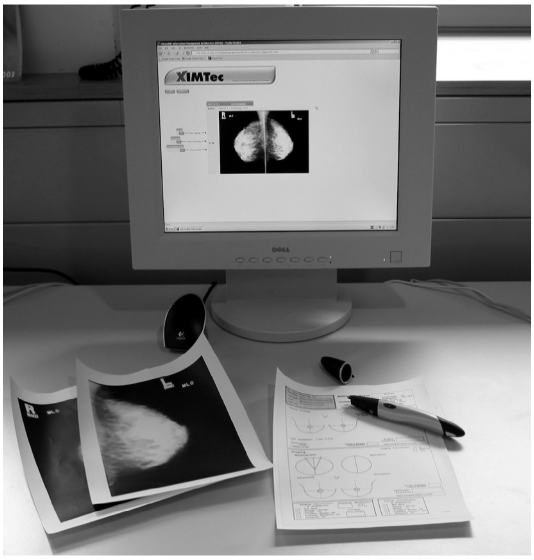

Fig. 6. Mammography reading workplace

One of the main problems has been the off-line mode of the Logitech pen. The whole creation of the annotations is done in a batch mode when the pen is placed in its cradle. This means that it is not possible to visualise any state or signal errors to the users during the process of annotation. This violates one of the basic principles of human-computer interaction [40]. Thus, interactivity is a must for the input device. As stated earlier, we believe that interactive versions of the pens will become available soon.

Existing annotations are available to the current users of the paper documents. The images corresponding to the annotations can be displayed on the computer screen along with additional information such as the creator of the annotation or the date. A smaller version of the mammogram is also displayed along with the active areas of the paper documents. However, in order to support *blind double reading*, i.e. independent analysis of the mammograms by different readers, annotations of other users are not visible by default. The readers perform their analysis and first create their annotations before consolidating their findings with those of the other readers.

The context-dependent visualisation of the annotations is one of the biggest advantages of our prototype over the traditional non-digital solution. In current work practices, multiple versions of the paper form have to be accessed to support blind double reading. Another advantage is the instant availability of the annotations in a digital form. There is no need to scan or transcribe the report forms.

In the current implementation, the system only supports unstructured annotations, i.e. hand-written texts or drawings. The only exception is a link to the patient record in the application database, which is anchored on the front page of the report form. Even though structured annotations are supported by the underlying framework, the authoring of such annotations is still under investigation. One of the limiting issues again is the missing on-line and interactive capabilities of the information capturing pens.

7 Conclusions

We have presented an extensible information infrastructure supporting various forms of cross-media annotation. Our framework is flexible in supporting new forms of resource types and also can be easily extended with new categories of annotations.

The annotation framework is based on a general link server that was developed within the Paper^{++}project for digitally augmented paper to support *enhanced reading*. However, since the Input Device API of this server handles the pen's position input in a device independent manner, it was straightforward to integrate Anoto pens as client input devices and therefore also support *enhanced writing*. By developing the appropriate authoring tools, we were able to develop an annotation framework that supports both the creation of and access to annotations across a range of media types, inclusive of paper.

The main limitations of the applications that we have developed for working with paper lie with the technologies and human interaction issues rather than the framework itself. The problems with existing readers arise because they were not developed with such applications in mind and they tend to focus on information capture rather than on interaction. However, the underlying technologies are general enough and it simply requires new configurations to be developed that support direct interaction. In terms of human interaction issues, the linking of the printed and digital worlds in this way opens up lots of new possibilities and problems to be investigated. The provision of general frameworks and prototypes such as ours is an important first step in facilitating these investigations.

Acknowledgements

This work is based on infrastructure developed as part of the European project Paper^{++}, under the Disappearing Computer Programme (IST-2000-26130). We thank the other members of the project for their contributions and feedback on our work. We also thank Rob Procter, Mark Hartswood, Roger Slack and James Soutter of the University of Edinburgh for valuable information and discussions on the use of annotation in breast screening.

References

1. Sellen, A. J., and Harper, R. *The Myth of the Paperless Office.* MIT Press, November 2001.

2. Marshall, C. C. Annotation: From Paper Books to Digital Library. In *Proceedings of DL'97, 2nd ACM International Conference on Digital Libraries* (Philadelphia, USA, July 1997).
3. Kahan, J., Koivunen, M.-R., Prud'Hommeaux, E., and Swick, R. R. Annotea: An Open RDF Infrastructure for Shared Web Annotations. In *Proceedings of WWW10, 10th International World Wide Web Conference* (Hong Kong, May 2001).
4. Geyer-Schulz, A., Koch, S., and Schneider, G. Virtual Notes: Annotations on the WWW for Learning Environments. In *Proceedings of AMCIS'99, 5th Americas Conference on Information Systems* (Milwaukee, USA, August 1999).
5. Resource Description Framework (RDF) Model and Syntax Specification, http://www.w3.org/TR/REC-rdf-syntax/.
6. Fensel, D., Horrocks, I., Harmelen, F. V., Decker, S., Erdmann, M., and Klein, M. OIL in a Nutshell. In *Proceedings of EKAW'2000, 12th International Conference on Knowledge Engineering and Knowledge Management* (Juan-les-Pins, France, October 2000).
7. Scientific Annotation Middleware (SAM), key = SAM, year = 2001.
8. Logitech io Personal Digital Pen, Logitech Inc., http://www.logitech.com.
9. Paper^{++} Project, IST-2000-26130, Disappearing Computer Initiative, http://www.paperplusplus.net.
10. Berners-Lee, T., Hendler, J., and Lassila, O. The Semantic Web. *Scientific American* (May 2001).
11. Gertz, M., Sattler, K.-U., Gorin, F., Hogarth, M., and Stone, J. Annotating Scientific Images: A Concept-based Approach. In *Proceedings of SSDBM'2002, 14th International Conference on Scientific Database Management* (Edinburgh, Scotland, July 2002).
12. Gertz, M., and Sattler, K.-U. Integrating Scientific Data through External, Concept-Based Annotations. In *Proceedings of DIWeb'02, 2nd International Workshop on Data Integration over the Web* (Toronto, Canada, May 2002).
13. Marshall, C. C. Toward an Ecology of Hypertext Annotation. In *Proceedings of ACM Hypertext'98, 9th Conference on Hypertext and Hypermedia* (Pittsburgh, USA, June 1998).
14. O'Hara, K., and Sellen, A. A Comparison of Reading Paper and On-Line Documents. In *Proceedings of ACM CHI'97, Conference on Human Factors in Computing Systems* (March 1997).
15. Hartswood, M., Procter, R., Rouncefield, M., and Slack, R. Performance Management in Breast Screening: A Case Study of Professional Vision and Ecologies of Practice. *Journal of Cognition, Technology and Work* (2001).
16. Want, R., and Russel, D. M. Ubiquitous Electronic Tagging. *IEEE Concurrency* (December 1999).
17. Rekimoto, J., and Ayatsuka, Y. CyberCode: Designing Augmented Reality Environments with Visual Tags. In *Proceedings of DARE'2000, Designing Augmented Reality Environments* (Elsinore, Denmark, April 2000).
18. Pradhan, S., Brignone, C., Cui, J.-H., McReynolds, A., and Smith, M. T. Websigns: Hyperlinking Physical Locations to the Web. *IEEE Computer* (August 2001).
19. Newman, W., and Wellner, P. A Desk Supporting Computer-Based Interaction with Paper Documents. In *Proceedings of ACM CHI'92, Conference on Human Factors in Computing Systems* (May 1992).
20. Wellner, P. Interacting with Paper on the DigitalDesk. *Communications of the ACM 36*, 7 (July 1993).

21. Koike, H., Sato, Y., and Kobayashi, Y. Integrating Paper and Digital Information on EnhancedDesk: A Method for Realtime Finger Tracking on an Augmented Desk System. *ACM Transactions on Computer-Human Interaction* (December 2001).
22. Aliakseyeu, D., and Martens, J.-B. Physical Paper as the User Interface for an Architectural Design Tool. In *Proceedings of INTERACT'2001, 8th Conference on Human-Computer Interaction* (Tokyo, Japan, July 2001).
23. Arai, T., Aust, D., and Hudson, S. PaperLink: A Technique for Hyperlinking from Real Paper to Electronic Content. In *Proceedings of ACM CHI'97, Conference on Human Factors in Computing Systems* (March 1997).
24. Hecht, D. L. Printed Embedded Data Graphical User Interfaces. *IEEE Computer* (March 2001).
25. Kafka, C. DataGlyphs Bridge Paper and Digital Worlds. *Docu World* (1998).
26. mimio Xi, http://www.mimio.com/meet/xi/.
27. Anoto AB, http://www.anoto.com.
28. Guimbretière, F. Paper Augmented Digital Documents. In *Proceedings of UIST'2003, 16th Annual ACM Symposium on User Interface Software and Technology* (Vancouver, Canada, November 2003).
29. Bechhofer, S., and Goble, C. Towards Annotation using DAML+OIL. In *Proceedings of Workshop on Knowledge Markup and Semantic Annotation* (Victoria B.C., Canada, October 2001).
30. Signer, B., Grossniklaus, M., and Norrie, M. C. Java Framework for Database-Centric Web Engineering. In *Proceedings of WebE'2001, 4th Workshop on Web Engineering (in conjunction with WWW10, 10th International World Wide Web Conference)* (Hong Kong, May 2001).
31. Millard, D., Moreau, L., Davis, H., and Reich, S. FOHM: A Fundamental Open Hypertext Model for Investigating Interoperability between Hypertext Domains. In *Proceedings of ACM Hypertext 2000, Conference on Hypertext and Hypermedia* (May 2000).
32. Norrie, M. C., and Signer, B. Switching over to Paper: A New Web Channel. In *Proceedings of WISE 2003, 4th International Conference on Web Information Systems Engineering* (Rome, Italy, December 2003).
33. Andrews, K. *Browsing, Building, and Beholding Cyberspace: New Approaches to the Navigation, Construction, and Visualisation of Hypermedia on the Internet.* PhD thesis, Graz University of Technology, 1996.
34. DeRose, S. J. XML Linking. *ACM Computing Surveys 31*, 4 (December 1999).
35. Grønbæk, K., Hem, J. A., Madsen, O. L., and Sloth, L. Designing Dexter-Based Cooperative Hypermedia Systems. In *Proceedings of ACM Hypertext'93, Conference on Hypertext and Hypermedia* (Seatle, USA, November 1993).
36. Halasz, F. G., and Schwartz, M. The Dexter Hypertext Reference Model. *Communications of the ACM 37*, 2 (1994).
37. Kobler, A., and Norrie, M. C. OMS Java: Lessons Learned from Building a Multi-Tier Object Management Framework. In *Proceedings of OOPSLA'99, Workshop on Java and Databases: Persistence Options* (Denver, USA, January 1999).
38. Norrie, M. C. An Extended Entity-Relationship Approach to Data Management in Object-Oriented Systems. In *Proceedings of ER'93, 12th International Conference on the Entity-Relationship Approach* (Arlington, USA, December 1993).
39. Hartswood, M., Procter, R., Rouncefield, M., Slack, R., Soutter, J., and Voss, A. 'Repairing' the Machine: A Case Study of the Evaluation of

Computer-Aided Breast Screening. In *Proceedings of ECSCW 2003, 8th European Conference on Computer Supported Collaborative Work* (Helsinki, Finland, September 2003), pp. 375–394.
40. Norman, D. A. *The Psychology of Everyday Things.* Basic Books, New York, 1988.