From ZigZag™ to BigBag: Seeing the wood and the trees in online archive finding aids

Ian G. Anderson
University of Glasgow
11 University Gardens
Glasgow G12 8QH
+44 (0)141 330 3843
I.Anderson@hatii.arts.gla.ac.uk

ABSTRACT
This paper reports on a one year speculative research project that sought to test the technical feasibility, practical implications and usability of transforming an XML Encoded Archival Description (EAD) finding aid into an XML ZigZag™ structure and applying a relational browser interface.

Categories and Subject Descriptors
H.5.4 [**Hypetext/Hypermedia**]: Navigation; User Issues

General Terms
Design, Experimentation, Human Factors.

Keywords
EAD, ZigZag™, Ted Nelson, XML, Browsing, Visualisation.

1. INTRODUCTION
On the whole the archive profession is a conservative and traditional one. Since its inception the principles of provenance, or *Respect des Fond,* and adherence to original order have been dominant characteristics in most archive communities. As a result the practice of describing archive collections in hierarchical arrangements is firmly embedded. Compared to other information services, however, standardisation, both in terms of descriptive standards and arrangement have been relatively late developments, as has the provision of online finding aids.

However, as more archival finding aids, of increasing complexity, become available online the difficulty of seeing the 'wood from the trees' increases. This is particularly the case when these are implemented in Encoded Archival Description (EAD) [1]. EAD is an XML DTD for the creation of machine readable, cross searchable archival finding aids and its creators consciously based its structure on hierarchical analogue finding aids. Whilst this provided an important comfort zone for archivists to migrate to encoded finding aids, it is also meant EAD inherited the innate difficulty of navigating hierarchical structures.

Whilst an archive's physical space, catalogue arrangement and archivist's assistance all help to guide users' navigation in the analogue world, this paradigm does not easily translate to the electronic. Nor has there been a significant body of research established on archive user's information seeking behaviour. Indeed there is little evidence that traditional archival arrangement adequately served the needs of users in the analogue world. It is unlikely, therefore, that replicating such arrangements in the digital world would prove any more successful.

Where research on archive user needs has been undertaken a range of characteristics have been discovered that suggest a more flexible approach to archival access is required. The very earliest studies in the late 1990s indicated that time, training and access to information about information were crucial barriers to electronic access, even though this access had become a critical component of historians' research methods [2]. Later studies have revealed the plurality of historians' information seeking behavior but also the need for both research and archival context that was common amongst the most popular methods [3] and the importance of intermediaries in the use of online material [4]. Academic historians require multiple pathways to access primary research materials and the need for user education on electronic searches suggests that current provision hinders access [5]. Moreover, the need for orientation in even the most experienced user has been emphasized [6].

Archive portal sites such as the Archives Hub, A2A, AIM25, ANW and SCAN are evidence of the desire to search across collections and repositories but typical means of browsing or displaying search results, such as lists and directories, severely restrict users' ability to see where they are, how they got there and where they can go next [7, 8, 9, 10, 11]. Providing linked 'cross-walks' such as subject keywords, functional descriptions, person, place and corporate names can only go so far in addressing this problem. Points at which these cross-walks intersect can not easily be displayed and users wishing to move from one to another need to repeat searches or navigate up and down the hierarchy.

This problem increases exponentially where related material is held in different series, collections or repositories. In these circumstances trying to follow a particular person, function or responsibility is extremely difficult. In following one path, users lose sight of others, where they cross and what their relationships are. In essence the multidimensional relationships that exist within the finding aid are subordinate to its hierarchical structure.

The threads of this research all combine to suggest that a means by which archive users can quickly and intuitively orientate themselves within collections and identify the relationships and context of the resources they are viewing would be immensely beneficial. This project, funded by the UK's Arts and Humanities Research Council (AHRC) through a one year speculative research grant, sought to test a novel approach to structuring and visualising archival information by applying a relational browsing interface to EAD finding aids that have been transformed into a multidimensional structure.

2. A MULTIDIMENSIONAL SOLUTION

One potential solution to this problem is to structure and visualise this information multidimensionally. For example, repository, collection, date and function could each be a separate dimension, rather like lines on a London Underground map. Therefore, a user viewing the person name dimension (or line) would see each individual represented in a finding aid as a cell. This person may appear in different parts of a collection, separate collections at the same repository and at other repositories, quite possibility related to different organisations, functions or roles. Whilst well developed finding aids can make these links, it is very difficult for users to see and navigate them.

One such means of organising information multidimensionally is the ZigZag™ concept developed by Ted Nelson [12]. In other words, a piece of information can exist in different places at the same time and have many connections to other information that may also exist in more than one place. The beauty of the ZigZag™ system is that the user can bring multiple instances of the same information into one view and by changing the dimensions can instantaneously see how the related bits of information are connected. Thus the user is always presented with a locally relevant view of the information, irrespective of how complex the structure is, and without losing the ability to navigate and view all the interconnections. The possibility to represent archival information in this way may provide both functionality and usability that reflects the deep interlinked structures of today's online finding aids. These additional dimensions could be used to provide a whole range of context specific information, such as related bibliographies, digital surrogates, user comments and help files. This would allow online finding aids to move from an access tool to an expert system.

The advent of XML encoded finding aids, particularly EAD, and the wide scale implementation of descriptive standards made this an ideal time to test the viability of a ZigZag™ structure and visualisation.

The number and extent of dimensions it is possible to represent, does of course, depend upon the quality and extent of the underlying data. For this project two finding aids, Gateway to Archives of Scottish Higher Education (GASHE) and Navigational Aids for the History of Science and Technology (NAHSTE), provided by the University of Glasgow Archive Services were selected [13, 14]. These finding aids provided the project with the opportunity to test the concept against EAD, the descriptive standards General International Standard Archival Description (ISAD(G)2) and International Standard Archival Authority Record for Corporate Bodies, Persons, and Families (ISAAR(CPF)). The GASHE finding aid also including function and activity 'cross walks' within it. Both finding aids cross multiple collections and repositories.

Overall, the project aimed to achieve 'proof of concept' status - that it was technically feasible to map between EAD and a ZigZag™ structure; that the transformation between the two could be automated; that a web based interface could represent the multidimensions and; that it supported more intuitive browsing for users.

3. DEVELOPMENT

Several working examples of ZigZag™ structures have already been created in other projects using Perl, C, Python and Java to run on Windows, Linux and Mac. Initially the most promising of these for this project was the combination of XML, XSL and JavaScript, successfully demonstrated by Les Carr at the IAM Research Group, University of Southampton on a map of the London Underground, see Figure 1 below [15].

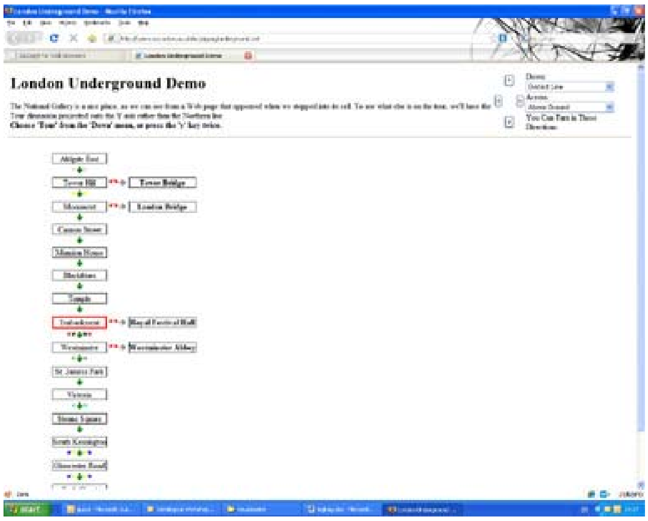

Figure 1. ZigZag™ for Web Browsers London Underground Demo

Taking this project as an inspiration and re-using the ZigZag™ for Web Browsers XML dialect kindly provided by Les Carr the project's functional and technical specification was defined using the Unified Modeling Language (UML), in particular use-case and activity diagrams. During this process it was decided that static, rather than 'on the fly', transformation of the finding aids were most appropriate in this context, given that the source data was static and the additional computational demands that online dynamic transformations would entail.

Mapping the EAD finding aids to the ZigZag™ structure through XSL was the first major project milestone. The finding aids actually comprise hundreds of files, in the case of GASHE, an XML EAD file using the ISAD(G)2 descriptive standard, XML ISAAR authority file, XML FANDA file (for functional and activity descriptions) for each individual collection, in each repository. The relationship between these files needed to be understood before the EAD file could be mapped to the ZigZag™ structure. At this point the application was named BigBag, partly as it was a nice alliteration of ZigZag™ but also because the dimension lines between cells in the mapping diagrams resembled the large string shopping bags used to carry groceries, see Figure 2 below.

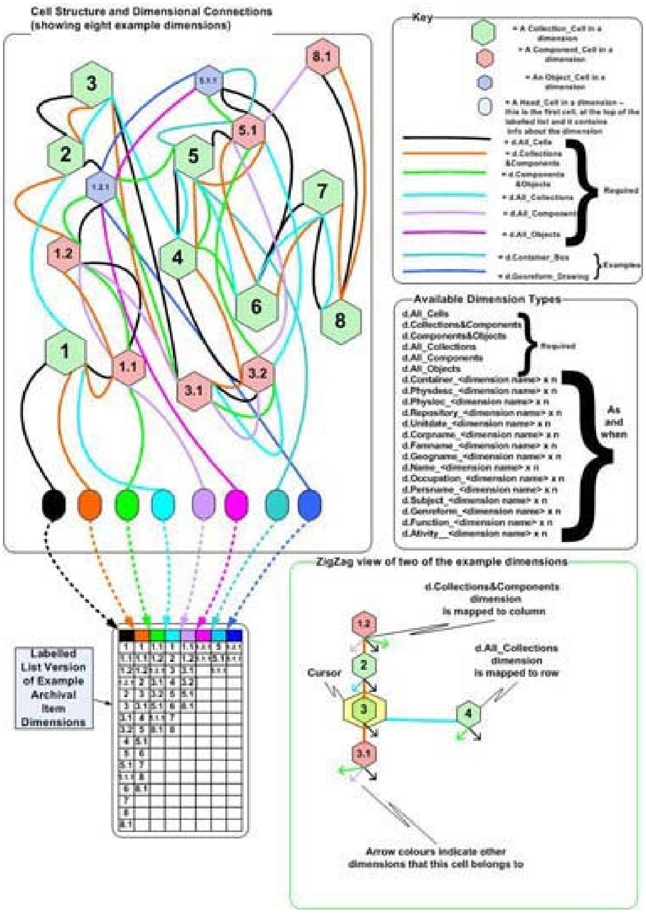

Figure 2. Archive Structure to ZigZag™ Structure Mapping

Thirteen EAD features were mapped to the ZigZag™ output tree by the projects 'Transform Finding Aid to ZigZag' stylesheet so each of the many ways to categorise an archival component became a dimension: subject, repository, personal name, location (shelf number etc.), container (box, folder, album etc.), format (book, film, letter etc.). function, date, century, business or corporate name, a daisy chain crossover linking all archive components, a crossover linking collection and series, and a crossover linking series and objects. As the underlying structure is one of linked circular lists the same cells may appear simultaneously in different orders in several linked lists.

The stylesheet had to manage some peculiarities of EAD and handle function, subject and place dimensions differently from other elements as multiple elements were possible, nested within a <p> tag. For example the stylesheet had to avoid adding the current place if it was the same as a place that had already been added, unless the current place was a sibling of the place that had already been added.

The stylesheet expected six required EAD elements, sixteen optional EAD elements and six optional multiple and recursive EAD elements. Seven escaped character codes were also stripped from the input tree as well as 13 characters that were illegal in JavaScript.

The cells of the output tree could be one of three types: collection (this included collection, fonds, class and record group descriptions), series (this included series, subfonds, subgroups or subseries descriptions) or object (this included item or file descriptions). Fragments from the XML ZigZag™ structure are provided in Figure 3 below.

```
<dimensions>
<dimension name="AllComponentsandObjects"
description="All Components and Objects" />
<dimension name="12thCentury"
description="12th Century" />
<dimension name="Subject:Accounting"
description="Subject: Accounting" />
<dimension name="Date:1971/1993"
description="Date: 1971/1993" />
<dimension name="Repository:Glasgow
Caledonian University Archives"
description="Repository: Glasgow Caledonian
University Archives" />
<dimension name="Format:file"
description="Format: file" />



<cells>
<cell n="1">
<url>http://www.gashe.ac.uk:443/cgi-
bin/view_isad.pl?id=GB-1847-
GP&amp;view=basic</url>
<title>Records of Glasgow Polytechnic
formerly Glasgow Collegeformerly Glasgow
College of Technology</title>
<content>fonds</content>
<link direction="AllArchiveComponents"
posward="2" />
<link
direction="AllCollectionsandComponents"
posward="2" />
<link direction="20thCentury"
posward="2"></link>
<link direction="Place:Cowcaddens Road
Glasgow" posward="2"></link>
<link direction="CorporateName:Glasgow
Polytechnic" posward="4"></link>
<link direction="Date:1971/1993"
posward="2"></link>
<link direction="Repository:Glasgow
Caledonian University Archives"
posward="2"></link>
</cell>
```

Figure 3. ZigZag XML Code

The transformation from EAD to ZigZag™ used Microsoft's Command Line Transformation Utility (MSXSL). This process was a two step transformation of archival finding aid data, from EAD XML into ZigZag™ XML and then into ZigZag™ HTML. Les Carr's ZigZag™ for Web Browsers is limited to 40 cells so a test file was selected that outputted 27 cells.

However, initial tests of a sample of data from the GASHE finding aid using Les Carr's XML dialect and JavaScript interface proved problematic. The transformation produced a functionally correct interface, but one that had limited usability, comprising hundreds of small black arrows dispersed across several screen widths, see Figure 4 below. Furthermore, even with the small sample data set, well specified PCs (dual core Pentium processors, 2GB RAM and 256MB dedicated graphics memory) were returning warnings that the JavaScript was causing the computer to run slowly. Although the number of cells was small, the number of dimensions associated with each cell in GASHE was far greater than in the original London Underground demo. These factors suggested that the JavaScript development path was unlikely to scale well enough for the amount of data and number of relationships required or provide sufficient complexity for the visualisation.

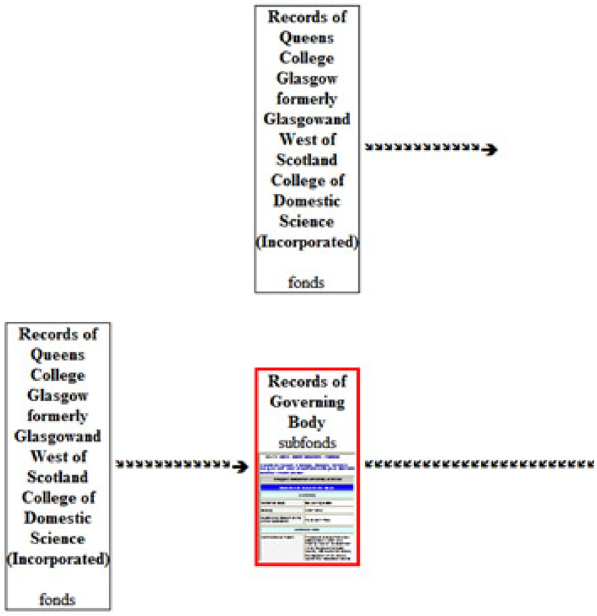

Figure 4. Section of BigBag JavaScript Demo

Whilst the appearance of the interface and the efficiency of the data handling could undoubtedly have been improved a decision was taken to seek an alternative means of visualisation. Initially an SVG interface was an attractive solution. It would keep the data within the XML family and the Parip Explorer project had successfully demonstrated a visualisation style that could suit the data [16]. However, the lack of project experience with SVG and the limitations of browser support led the project, after further research, to develop its interface using Macromedia/Adobe Flash based on an original idea by Moritz Stefaner [17]. Stefaner's relational browser for the CIA World Fact Book provided the underlying physics for an interface that positioned the selected 'cell' in the centre of the screen with lines spanning out to related cells of information. Selecting an outlying cell brought this to the centre of the screen and redrew the relationships. In other words it provided users with locally relevant view of their selected information without losing sight of the immediately bigger picture. An initial trial with a simple greyscale version of the relational browser interface demonstrated that it was capable of being modified to reflect, in part at least, the underlying ZigZag™ structure.

The second version of the interface, and the first to be tested with users, added a colour keyed sliding selector for the various dimensions as well as drop down menus for selecting instances of dimensions and archive components. A breakout box that linked to the original finding aid for each selected cell was also added as well as history and home buttons, see Figure 5 below [18].

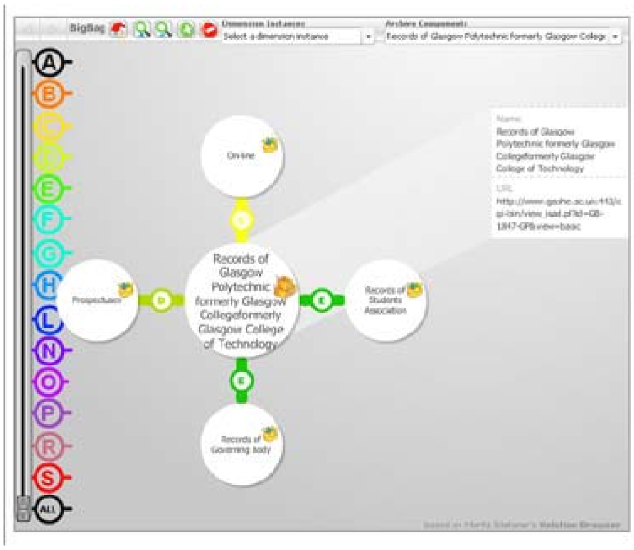

Figure 5. BigBag Flash Demo Version 2

A small, targeted sample of six people, two archivists, two historians and two students were selected to test this first version. Although the feedback was positive on the whole, with participants finding the interface clear, intuitive and supporting their browsing behaviour it was also evident that the multidimensionality of the underlying ZigZag™ structure was not being adequately expressed. Stefaner's relational browser only had to express one type of 'part of' relationship between two cells at a time and employed a single line to do so. However, with the finding aid ZigZag™ data there are potentially many different relationships between each cell that a single line cannot adequately convey. The sliding dimension selector was an attempt to overcome this problem but users did not like having to scroll through each dimension on the slider to see if it applied to their selected cells. It was evident that a means of immediately representing the number and type of relationships between cells was needed.

The next version of the interface, version three, tested the technical possibility of having multiple lines, each representing a different dimension, connect each cell and for the width of these lines to reflect the number of instances within that relationship. Once the project had established that this was technically and aesthetically possible version four of the interface was released. This removed the sliding dimension selector and replaced it with a simple key to the coloured lines. The format of the breakout box to link to the original finding aid was simplified and the screen split to show the original finding aid to the right [19]. In this version of the visualisation the colour of the line again indicates the dimension type with line widths indicating the number of instances for each type, the thicker the line, the greater the number of instances. Icons indicated whether the cell data existed at the collection, series or object level in the original finding aid, see Figure 6 below.

Figure 6. BigBag Flash Demo Version 4

The same six users who previously evaluated the project were shown the final version of the interface. In this case the underlying relationships within the data were agreed to be more explicit and enhanced the browsing of the finding aid data. However, users now instinctively wanted to click on the connecting lines to isolate a particular dimension, a functionality that was not possible, rather than use the dimension instance drop down menu. Furthermore, bugs and inconsistencies, particularly in the way dimension instances were selected significantly hampered users. Selecting a subject dimension form the drop down menu did not alter the cell display, an error that was not present in the previous version of the interface, and an additional erroneous subject instance also appeared on the menu.

By this late stage in the project time was a major constraint and it was not possible to either complete the scheduled interface testing or implement the zoom in and out function, or the add and subtract cells feature. Indeed it was a struggle to get the final version of the visualisation working in time at all.

4. STRENGTHS & WEAKNESSES

In part the project has fulfilled its main objectives. It established that it was conceptually possible to map from EAD to ZigZag™ and that a stylesheet could be developed that automated this process. However, it was not possible to establish that this transformation could be undertaken on all instances of EAD finding aids. Even working within the GASHE collection, variations in EAD encoding practices posed a challenge to efficient transformations. In part this is an inherent weakness of EAD in that its minimal compliance requirement amounts to little more than a collection description, akin to a minimally compliant TEI header. In the projects test data the lack of entity declarations for special characters also interfered with attempts to create suitable visualisations. In retrospect, editing the GASHE EAD prior to transformation would have created a far more efficient process. However, in trying to create a transformation that would be applicable to real life situations it would be unrealistic to expect archivists to amend their EAD files in order to accommodate our visualization. One useful spin-off from this, however, was the development of a set of EAD templates for the NoteTab text editor that placed greater constraints on coding choices. The objective of this exercise was that archivists might adopt them when creating new EAD finding aids and so avoid many of the common problems found in the EAD finding aids that hindered this project.

Although the project was able demonstrate the technical viability of an XML ZigZag™ for web browsers on larger and more complex data than the London Underground demo, this was not significantly so and time did not allow for the transformation and visualisation of the entire GASHE finding aid let alone test the stylesheet against NAHSTE.

Throughout the project a difficult balance had to be struck between refining and testing the stylesheet against larger and more varied sets of source data and developing a meaningful visualisation to test with users. In the end neither component was as fully developed as it could have been, but the project would have failed in an important respect if it had successfully transformed a large amount of data without any means of displaying the results. In retrospect the project may simply have been too ambitious in its scope.

After a few false starts the project did create a visualisation that reflected the underlying multidimensionality of the ZigZag™ data, albeit imperfectly. Although the fourth and final interface is the closest conceptually to the goals the project set itself its limited development time, even compared to the second version, proved a hindrance to establishing with certainty that this provided a significantly more beneficial interface to online archive users. It was never the projects intention to undertake extensive user evaluation or usability testing but within the constraints of what was possible the generally positive feedback is sufficient to suggest that the approach adopted does bring benefits for browsing archive finding aids online. How great those benefits are, for what type of information seeking behaviour and in what circumstances are questions that this project is unable to answer.

5. CONCLUSION

Perhaps inevitably for speculative research this project ultimately raises more questions than it answers, but has at least demonstrated sufficient merit to warrant those questions being investigated further. In particular the relative importance of the underlying EAD finding aid, ZigZag™ structure and visualisation on the end user's understanding of the data needs to be examined.

It is the intention to continue this research by creating a set of alternative structures and visualizations based on the same underlying archive data – a relational visualisation directly on an EAD fining aid; archive data that has been directly inputted to a ZigZag™ structure rather transformed; an EAD to ZigZag™ transformed visualisation (essentially an updated version of the current visualization) and; the archive data as displayed in its native state.

These alternative representations will provide a test bed through which end users understanding of the archive data will be examined using reception theory. Reception theory, sometimes called audience response theory, is a version of reader response theory that first developed in literary studies and was subsequently extended to include performance works. Reception theory proposes that a text does not have an inherent meaning, but meaning is created within the relationship between the text and the reader, shaped by the reader's background, influences and biases. By applying this theory to archival data it is hoped to explore the extent to which meaning is created by the user, is inherent to some extent in the data itself, and/or meaning is shaped by the way in which the data is structured or visualized.

There is also the potential for the approach tested here to be applied to information domains other than archives. Since this research was completed a brief market analysis was conducted to try and identify other areas that might benefit from this approach. Although this survey was by no means comprehensive, and there are a range of commercial data visualization products already available, the areas of social networks, personal or business contact lists, customer relationship management and enterprise relationship management are potentially new areas that future research could address.

6. ACKNOWLEDGMENTS

My thanks to the AHRC for funding, the Humanities Advanced Technology and Information Institute and the Faculty of Arts at the University of Glasgow for additional support, Lesley Richmond and Victoria Peters of University of Glasgow Archive Services for the supply of data and advice, Les Carr at the University of Southampton for permission to re-use his XML ZigZag™ for Web Browsers dialect, Moritz Stefaner for permission to re-use the underlying design of his relational browser and Steve North, Research Assistant, for his creativity, dedication and hard work.

7. REFERENCES

[1] EAD http://www.loc.gov/ead/

[2] Andersen, D.L. "Academic Historians, Electronic Information Access Technologies, and the World Wide Web: A Longitudinal Study of Factors Affecting Use and Barriers to that Use", The Journal of the American Association for History and Computing 1, June 1998. Available at: http://mcel.pacificu.edu/jahc/1998/issue1/articles/andersen/

[3] Anderson, I. "Are you being served? Historians and the Search for Primary Sources", Archivaria, 58, Fall 2004.

[4] Duff, W. Craig, B. Cherry, J. "Historians Use of Archival Sources: Promises and Pitfalls of the Digital Age," The Public Historian 26, no 2, Spring 2004.

[5] Tibbo, H. "Primarily History: How US Historians Search for Primary Sources at the Dawn of the Digital Age," American Archivist 66, no 1. Spring/Summer 2003.

[6] Duff, W. and Johnson, C. "Accidentally Found on Purpose: Information Seeking Behaviour of Historians in Archives," Library Quarterly 72, no. 4, 2002.

[7] Archives Hub http://www.archiveshub.ac.uk/

[8] Access to Archives (A2A) http://www.nationalarchives.gov.uk/a2a/

[9] AIM25 http://www.aim25.ac.uk/

[10] Archives Network Wales (ANW) http://www.archivesnetworkwales.info/

[11] Scottish archives Network http://www.scan.org.uk/

[12] ZigZag http://www.xanadu.com/zigzag/

[13] GASHE http://www.gashe.ac.uk/

[14] NAHSTE http://www.nahste.ac.uk/

[15] Les Carr's Zigzag for Web Browsers London Underground demo http://users.ecs.soton.ac.uk/lac/zigzag/

[16] Parip Explorer Project http://parip.ilrt.org/

[17] Moritz Stefaner's Web Site http://der-mo.net/

[18] BigBag Interfacev2 http://www.hatii.arts.gla.ac.uk/research/visual/demov2/index.html

[19] BigBag Interfacev4 http://www.hatii.arts.gla.ac.uk/research/visual/demov4/index.html