1. INTRODUCTION

One area of human-computer interaction (HCI) that receives a lot of attention, but from many different perspectives, is the design of coordinated multiple views (CMV) of data. CMV\(^1\) covers a vast range of different forms of interaction from scientific visualisations, to programming environments to ‘dash boards’ showing analysis data such as Google Analytics. But with such a vast range of possibilities, what is the appropriate choice of CMV for a given activity?

Put simply, the advice to creators and users of CMVs is to "define your environment around your tasks". This might, for example, require writing bespoke SQL queries and selecting interface widgets to present the results in a given view. But with the advent of "big data" on the one hand, and open-ended exploration of that data on the other, plainly the "task" can no longer be well defined. Organisational agility might require detection of novel patterns and insights into past data. For security and performance reasons, and perhaps due to lack of available skills, the cost of specifying and writing a bespoke query might be prohibitive, and the implementation time lag might mean the information is useless by the time it emerges.

Good database design, using well established techniques, ensures that the tables of data are organised in such a way that the known relationships are easily and efficiently accessed and modelled. What we are interested in, however, is when unknown relationships only emerge when data about the data (ie "metadata") reveals previously unknown linkages, or can answer questions that reflect unanticipated scenarios. Folksonomies are one example of such post hoc categorisation. We could go even further to help the user model the rate of change within dynamic visualisations, and thus predict future relationships. This has led us to seek to establish guidelines (or heuristics) for the design and use of CMVs to cover these new patterns of use.

In this paper, after reviews of data analytics, and of the notions of navigation of information space, we introduce a case study of a management information system (MIS) for a research institute, and show how people are able to navigate the data set (the ‘information space’) more effectively, if they have access, through a CMV, to metadata. From this study we derive an initial set of design heuristics which we plan to evaluate and develop in future studies, and identify some requirements for the next round of studies.

LITERATURE REVIEW

In HCI, CMV dates back to the late 1990s, when a CHI workshop led to a collection of papers in the Advanced Visual Interfaces (AVI) conference in 2000. Balonado, Woodfruss, & Kuchinsky, (2000) defined a CMV as a visualisation where more than one view was provided to ‘support the investigation of a single conceptual entity’ and, even then, pointed out that many different forms and many different systems existed in many different domains. North & Shneiderman, (1999) offered a high level taxonomy of CMVs based on whether the focus was on selecting items or navigating views, and whether the

\(^{1}\) Amongst the multiple perspectives is that, within the literature, CMV is sometimes stated as MCV, and with "Visualisations" in both the UK and US spellings replacing "Views".

© The Authors. Published by BCS Learning and Development Ltd.
Proceedings of HCI 2014, Southport, UK.
different views represented the same or different information. They highlighted three significant features of a CMV: benefits to user performance, discovery of novel relationships and the management of complex information through interaction.

At the same conference they (North & Shneiderman, 2000) suggested that the relational data model (Codd, 1970) provided a good basis for coordination, equating the object of a visualisation with a tuple of a relation. Different CMVs could be automatically ‘snapped together’ based on the relationships between relations as either one-to-one, or one-to-many. Boukhelifa, Roberts, & Rodgers (2003) however, argue that just using the relational model is too restrictive. They go on to develop their view of coordination focusing on how the different views are coordinated.

The Information Visualisation community (InfoVis) has also focused on CMVs, typically coming at CMV from the perspective of drawing graphs, avoiding crossing lines and other automatic methods for generating attractive visualisations.

Despite the breadth of work in this area, there is surprisingly little discussion about which data should be presented in the visualisation. It is generally assumed that the visualisation designer will discover the tasks that users are trying to do and will design a CMV to support those tasks (Baldonado et al., 2000). There are many examples of this, but little generalisation. Chen (2005), summarises these issues, identifying in particular a lack of evaluative methodologies for usability in CMVs, a need to shift design thinking from structural to dynamic, and an overall lack of robust empirical work in the area. We seek to rectify this by proposing simple, but powerful design heuristics for a particular class of CMVs and evaluating these through adequate empirical studies.

Stated simply, our initial design heuristic is ‘providing data and metadata for people using a CMV will improve their ability to navigate a data set’.

**Navigating MIS via metadata**

The interest of this research is to design CMVs to afford the different aspects of navigation which Benyon & Höök (1997) characterised as object identification, wayfinding and exploration, and what Spence (2002) terms the process where people gain knowledge of an information space, and are able to interpret this knowledge effectively in order to retrieve adequate information.

We focus here on the use of CMVs for MIS, rather than scientific data, geographical data or pictorial databases. We are interested in the analysis of management data and in developing CMVs to support visual analytics: ‘the science of reasoning facilitated by interactive visual interfaces’ (Thomas & Cook, 2005). These authors present a research agenda that was developed in response to the terrorist attacks in the USA in September 2001. It is a hugely ambitious and comprehensive attempt to create a new science of analytical reasoning and the book covers theories of reasoning, representations, cognition and interaction at the many different levels of detail that establishing such a new science requires.

We wish to apply these ideas in the use of CMVs in MIS. Such situations are extremely common, whether the aim is to manage a research institute, a service industry (such as “care in the community” services), or a garden centre. Rather than the data retrieval focus of traditional database use, analytics focuses on trends, seeing new relationships between objects and navigating the database in new and interesting ways.

In his state-of-the-art review of CMVs, (Roberts, 2007) discusses “view generation” as an important part of the development of CMVs: designers need to consider the form of the visualisation, how to map information to the form, how to abstract and aggregate data, and how the user interacts with the data. However, he does not consider what data to present in the first place.

We think the issue of how to aggregate the data requires designers to consider the development and use of metadata, which traditionally is seen as a different type of data from the underlying database. For example in web design, (Wodtke, 2003) identifies three types of metadata: *Intrinsic* (file size, the resolution of graphical images and so on), *Administrative* (author’s name, date-created, and so on) and *Descriptive* (highlighting useful facets of an object). In relational database design, data dictionaries are the place for metadata (Elmasri & Navathe, 2007).

However, hash tags, as used in social media, are also metadata (and typically *Descriptive*). A picture on Instagram, for example, will have a hash tag such as ‘sunset’ allowing people to search and find all the pictures of sunsets. Combining with other tags such as the location ‘Boracay’ will find pictures of sunsets in Boracay. So, although this is classified as metadata, there is no intrinsic difference between the data and the metadata. The data about the picture (sunset in Boracay) is the metadata that will help people find the picture they want. So one person’s data is another person’s metadata. Most recently this distinction has been made in the context of the National Security Agency (NSA) in the USA accessing the metadata about e-mails (Landau, 2014). This includes the e-mail addressee, where it came from, the subject, and information concerning when it was sent and received. Thus metadata about an e-mail becomes very useful data for the analytics in which the USA’s NSA is engaged. Plainly patterns of metadata allow inference and
deduction, or at least identification of data for further scrutiny with an enhanced probability of relevance.

There are two key problems that arise for the design of visual analytics. Baldonado et al say that designers ‘necessarily’ begin by establishing a clear view of the user’s task, when, of course, different people will have very different views, and hence different tasks, on any data set. Secondly, equating a tuple of a relational database with an object in the data set mixes up the data and metadata. A relation provides metadata for the object that is the primary key of the relation.

The issue of user tasks is critical. Shneiderman & Aris (2006) list a number of tasks that users of visualisations will want to do, such as ‘count the number of nodes and links’, ‘count the degree (the number of links) for each node’, ‘find the distance from one node to another (count the number of steps from one node to another)’, before finally concluding that ‘there are an unlimited number of tasks that could be defined’. Collins & Carpendale (2007) discuss the relationships between data, relations and visualisations, giving CMVs as one example of a multiple view. Step 1 of their method for creating visualisations is ‘choose a relationship’, without saying which relationships might be useful to choose. Javed & Elmqvist (2012), in their review of choosing and using multiple, coordinated views of a data set, look for recurring design patterns. They discuss the importance of 1:1 and 1:M links between relations but, again, stop short of saying which relationships might be useful to choose.

A typical CMV representation of this data is shown in Figure 1 (which is presented in low resolution so as not to identify individuals). This is a real example where the designer has taken a number of decisions about what to represent and how to represent it. The CMV contains four areas, which we will refer to as A (top left), B (bottom left), C (top right) D (bottom right).

Members of the institute are displayed in A, colour-coded by centre, with lines showing the co-published relationship (line thickness indicates volume of co-publications). B’s word-cloud displays key words, with size indicating frequency. C’s bar chart describes publications by year (X Axis), and type (colour coded). D is a matrix of people and shows the co-publishing relationship. Clicking a node in A, for example, selects the appropriate keywords for the individual in B, highlights their co-publications in D and individual details in C. These are typical areas of functionality implemented based upon an understanding of known needs of users at the time of implementation.

The CMV shown in Figure 1 supports a number of user tasks, but demonstrates some clear design flaws. The main problem with the design is that there is no key to the coding scheme used to encode centres or to know that the lines in the top left pane indicate the Co-published relationship and that the thickness of the line represents the number of co-publications. This may be intuitive to some but not to all. Thus the first piece of metadata that must be included with any visualisation is a description of the coding schemes used. The top right pane (C) does include this, but there is no description of the meaning of the y axis. There is no key either to the pane B or to colour coding in pane D.

Our design approach is to bring the concepts and design guidance from navigation in the real world to navigation in information spaces of data sets. We aim to design based on understanding data alongside its metadata, and on understanding how humans navigate in an information space.

RESEARCH OBJECTIVE

For example, a typical research institute keeps data about the people in the institute, the centre (or department) to which they are attached, the publications they have made, and the grants that they have obtained. Often there are collaborations across centres within the institute, members’ roles (and affiliations to a centre) change over time, as do areas of interest in research. In addition, each person, centre, grant and publication is associated with a keyword. A typical relational database implementation of this would consist of the following relations (with the primary key underlined):

- **Person** (PersonID, Name, CentreID)
- **Centre** (CentreID, Name)
- **Publication** (PublicationID, type, location of publication, year)
- **Grant** (GrantID, Amount, Funding Body)
- **Co-Published** (PersonID, PersonID, PublicationID, keyword)
- **Co-Investigated** (PersonID, PersonID, GrantID, keyword)

The relations Person, Centre, Publication and Grant record details of these objects. Co-Published represents the relationship between the people who co-publish a publication. Co-Investigated records the relationship between people who have collaborated on a grant.

A typical CMV representation of this data is shown in Figure 1 which is presented in low resolution so as not to identify individuals. This is a real example where the designer has taken a number of decisions about what to represent and how to represent it. The CMV contains four areas, which we will refer to as A (top left), B (bottom left), C (top right) D (bottom right).

- There are two key problems that arise for the design of visual analytics. Baldonado et al say that designers ‘necessarily’ begin by establishing a clear view of the user’s task, when, of course, different people will have very different views, and hence different tasks, on any data set. Secondly, equating a tuple of a relational database with an object in the data set mixes up the data and metadata. A relation provides metadata for the object that is the primary key of the relation.

- The issue of user tasks is critical. Shneiderman & Aris (2006) list a number of tasks that users of visualisations will want to do, such as ‘count the number of nodes and links’, ‘count the degree (the number of links) for each node’, ‘find the distance from one node to another (count the number of steps from one node to another)’, before finally concluding that ‘there are an unlimited number of tasks that could be defined’. Collins & Carpendale (2007) discuss the relationships between data, relations and visualisations, giving CMVs as one example of a multiple view. Step 1 of their method for creating visualisations is ‘choose a relationship’, without saying which relationships might be useful to choose.

- Javed & Elmqvist (2012), in their review of choosing and using multiple, coordinated views of a data set, look for recurring design patterns. They discuss the importance of 1:1 and 1:M links between relations but, again, stop short of saying which relationships might be useful to choose.

- A typical CMV representation of this data is shown in Figure 1 (which is presented in low resolution so as not to identify individuals). This is a real example where the designer has taken a number of decisions about what to represent and how to represent it. The CMV contains four areas, which we will refer to as A (top left), B (bottom left), C (top right) D (bottom right).

- Members of the institute are displayed in A, colour-coded by centre, with lines showing the co-published relationship (line thickness indicates volume of co-publications). B’s word-cloud displays key words, with size indicating frequency. C’s bar chart describes publications by year (X Axis), and type (colour coded). D is a matrix of people and shows the co-publishing relationship. Clicking a node in A, for example, selects the appropriate keywords for the individual in B, highlights their co-publications in D and individual details in C. These are typical areas of functionality implemented based upon an understanding of known needs of users at the time of implementation.

- The CMV shown in Figure 1 supports a number of user tasks, but demonstrates some clear design flaws. The main problem with the design is that there is no key to the coding scheme used to encode centres or to know that the lines in the top left pane indicate the Co-published relationship and that the thickness of the line represents the number of co-publications. This may be intuitive to some but not to all. Thus the first piece of metadata that must be included with any visualisation is a description of the coding schemes used. The top right pane (C) does include this, but there is no description of the meaning of the y axis. There is no key either to the pane B or to colour coding in pane D. Again, past user experience or training might make this
Designing Coordinated Multiple Views of Information Space
McEwan • Igonderigha • Benyon

understood, but if, say, a blogging platform changes the nature of how users interface with, say, a word-cloud, then the interaction breaks down. We therefore think it is reasonable to state the following design guideline for a visualisation: include the ‘first level’ metadata of an object, in the form of a key, a description of the coding scheme.

The visualisation supports certain tasks quite well such as ‘how many publications has person X made’, ‘who has person Y co-published with’, ‘when has person Y published,’ etc. However, the interface does not support tasks such as ‘how many publications have been made by Centre A’, ‘how many journal publications are there’ or ‘which keywords are associated with Centre B’, ‘which centre has had the greatest growth in publications per centre member in the last three years’. For each of these queries, the user has to aggregate the data by hand across the individual entries.
Data about centres, publications and grants is not easy to find using the visualisation in Figure 1. As soon as the user starts asking questions about the metadata, the visualisation breaks down. This situation can be solved with respect to data about centres as illustrated in Figure 2 by introducing a visualisation of the metadata of centres. In this case the designer has replaced the bar chart in the top right pane with a matrix representation of the centre collaborations (C1) and a network representation (C2) of the same relationship.

The visualisation still suffers from the lack of a key, i.e. the metadata about the coding scheme. Data about grants and publications could be provided with similar visualisations taking the grant relation or the publication relation as central. Clearly providing summary data as the metadata makes it easier to answer questions at that level.

However, now that centre directors are able to access the data they need, pretty soon the institute director will want to ask questions about the institute as a whole. Someone else will want to analyse the different journals or conferences that have been the locations of publications. The institute director might be asked to extrapolate future targets from past performance. So the analysis moves up another level of metadata which again is represented by the M:1 relationships and further relations in the database.

Thus we decide to evaluate the two CMVs with a large group of novice users.

The Design Approach

If we return to the general issue of how to design CMVs, we can see that it is necessary to allow people to move from the view of data to a view of metadata, and back again, depending on the focus of their queries. This involves a move from providing the CMV of a relation (or interrelated relations at the same level) to a visualisation that provides ways to move between data and metadata.

Our approach does not focus on user tasks, as tasks do not provide the flexibility that is provided by the navigation of information space approach (Benyon, 1992). Instead the designer needs to decide on the main focus of the data set and on the key relationships based around that focus. Then the designer should provide access to the metadata of that focus, both in terms of a key to the visualisation and in terms of the attributes of the many-to-one relationships.

EVALUATION

All first year students (n=165) of Computing across a wide range of different programmes were invited to participate in the study as part of a timetabled two hour class in HCI. We refer to these classes as “tutorials” although each has 30-48 students, normally working in groups of 6-8. Earlier that week, at the lecture for the full class, the research supervisors presented a series of slides explaining the research institute CMV, the aims and objectives of the research and an overview of what to expect during the experiment. The students were given time to ask questions and to get to meet the researcher. We advised the students that this tutorial had no bearing on their assessment, but that it was a chance for them to use some state-of-the-art equipment, and to get an insight into the university’s research.

The lab used for the study allowed up to three “sets” (a small group of 1-4) to work in parallel at large touch-screens, and the room could accommodate up to 12 students at a time. Students were timetabled 8 at a time into 20 minute timeslots, of which 10 minutes was given over to general teaching activity – information about the lab and the nature of research, and to read and sign informed consent forms. The remainder of the time was available for students to carry out a number of predefined tasks using one of the two CMVs, and complete a short feedback sheet. In the end 101 students turned up, and from these n=92 cooperated. The participants estimated their internet experience as being 4 to 10 years, while their daily internet usage ranged from 4 to 8 hours on weekdays and about 5 to 10 hours at weekends. They were all first time users of the CMV of the research institute. 46 participants used the CMV in Figure 1, 46 that in Figure 2.

The purpose of this initial empirical study was to lay the groundwork for future studies by examining the extent to which the two CMVs supports users to carry out the three types of navigation activity: object identification, exploration and wayfinding (Benyon and Höök, 1997). Each of the tasks selected for the students to carry out addresses one or more of these activities, while at the same time being typical of MIS tasks and relatively easy for the students to understand (see Table 1).

The objective of this experiment was to test the research hypothesis that providing data along with its metadata (in this case summary data about centres) in a CMV will make it easier and quicker for people to move around the information space. This should lead to successful navigation and hence improved information retrieval.

We also gathered data on the actions people take (using a screen logger), why they take them (optional post-session interviews) and the time it takes to complete a task (key-logging). The experiment adopted a between-subjects approach.

Experiment Setup
There were four “tutorials” in all, with 18, 37, 17, and 30 participants attending each (this represented an attendance rate of 60-65% in each tutorial, a typical turnout). Participants in the first session were given a sheet listing 10 tasks with space provided for answers to be written. This was more than could be accomplished in the time frame and this was primarily to help us establish a realistic task list for the subsequent tutorials. Each session lasted about 20 minutes, 10 minutes for participants to familiarize themselves with the system and 10 minutes to perform the tasks. The researcher observed the participants as they worked, made notes of comments, questions and interesting events.

Informal interviews were conducted, after the session, with participants who showed further interest, 40 in total, 10 from each tutorial group. Overall, we aimed to capture the following:

(i) The task type (object identification, exploration and wayfinding).
(ii) The number of mouse clicks involved.
(iii) The time taken to complete each task.
(iv) The extent to which features provided support successful navigation.
(v) Ability to navigate to specific information.
(vi) Ability to find relationships and spot trends.
(vii) User satisfaction.

Table 1: Task categorisation for the study

<table>
<thead>
<tr>
<th>Task</th>
<th>Group 1 (n=18), Interface A</th>
<th>Group 2 (30) Interface A</th>
<th>Group 3 (16) Interface B</th>
<th>Group 4 (30) Interface B</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Identify and list two different things on the visualisation and write down the information they provide</td>
<td>Object Identification</td>
<td>The goal is to see how participants can effectively identify and categorize groups and clusters of objects spread across the given interface. It will also show how well they understand the organisation of these objects and if they can retrieve adequate information about them.</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>How many publications has person X made?</td>
<td>Wayfinding and Exploration</td>
<td>The wayfinding aspect of the task demonstrates how well participants are able to interpret the interface, observe events in the different displays and find their way around the interface as they search for specific information. The exploratory part shows their ability to find relationships that exist, in this case, between an individual and their publication.</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Locate two people, in the Centre for Interaction Design (CID), who have made joint publications with people from other centres?</td>
<td>Exploration</td>
<td>This task aims to examine participants’ ability to find relationships that exist within the surrounding (for example, centres) and other surrounding (centres) and locate common patterns (publications) and trends that exist.</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>List the 5 Centres in the institute and identify the centre with the least number of members.</td>
<td>Object Identification and Wayfinding</td>
<td>This task shows the level of information participants can retrieve about the categorisation and organisation of a group of objects.</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>What subject is the Centre for Emergent Computing (CEC) interested in?</td>
<td>Object Identification, Wayfinding and Exploration</td>
<td>This task will point out participants’ ability to retrieve adequate information about a centre. Identify and locate the centre, then discover the relationship between the centre and its subject of interest.</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>How many grants have been awarded to the Centre for Social Informatics (CSI)?</td>
<td>Wayfinding and Exploration</td>
<td>The wayfinding phase of this task will show participants’ ability to reach a destination (find a centre) while the exploratory part will demonstrate their understanding of existing relationships.</td>
<td></td>
</tr>
</tbody>
</table>

Results

Table 2: Results: Task Completions

<table>
<thead>
<tr>
<th>Task</th>
<th>Group 1 (n=18), Interface A</th>
<th>Group 2 (30) Interface A</th>
<th>Group 3 (16) Interface B</th>
<th>Group 4 (30) Interface B</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>14</td>
<td>25</td>
<td>16</td>
<td>22</td>
</tr>
<tr>
<td>2</td>
<td>12</td>
<td>23</td>
<td>16</td>
<td>30</td>
</tr>
<tr>
<td>3</td>
<td>11</td>
<td>19</td>
<td>16</td>
<td>28</td>
</tr>
<tr>
<td>4</td>
<td>No data</td>
<td>13</td>
<td>16</td>
<td>23</td>
</tr>
<tr>
<td>5</td>
<td>8</td>
<td>2</td>
<td>13</td>
<td>24</td>
</tr>
<tr>
<td>6</td>
<td>6</td>
<td>2</td>
<td>13</td>
<td>22</td>
</tr>
</tbody>
</table>

Clearly the revised interface (Interface B) supports the tasks better than the original (Interface A). The revised interface resulted in completion of around 50% more tasks, whether measuring by individuals or by sets. The two smaller tutorials had higher task completion statistics than the larger tutorials. While we were able to ensure then ten minutes per test session was strictly adhered to, this variation may be due to a more hectic throughput of students in the two larger tutorials, which perhaps left a number of distractions and/or slightly disoriented the subjects.
Interestingly, the number of students in a Set does not appear to affect task completion rate:

<table>
<thead>
<tr>
<th>set size</th>
<th>tasks per set size</th>
<th># of set of size</th>
<th>average</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>49</td>
<td>12</td>
<td>4.1</td>
</tr>
<tr>
<td>2</td>
<td>98</td>
<td>25</td>
<td>3.9</td>
</tr>
<tr>
<td>3</td>
<td>48</td>
<td>12</td>
<td>4.0</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>1</td>
<td>4.0</td>
</tr>
</tbody>
</table>

### FUTURE WORK

This paper reports the initial results from our study, which we find encouraging. Regardless of the number of people working together in a small set, a similar number of tasks can be completed, but the CMV which supports metadata appears to improve task completion rates, compared to the CMV which does not.

A subsequent study was undertaken using a within-subjects approach that further investigated the hypothesis “when metadata is added to data in a CMV, users can navigate more effectively”. This extended the study reported here, by providing an interface that allowed people to select the metadata displayed in section C of the CMV. They could obtain summary data for grants, centres, publications or keywords.

The results confirmed the positive effect of providing metadata to assist in the navigation of the class of CMVs that we were investigating, namely those aimed at management information.

### CONCLUSIONS

Currently, CMV users find it challenging to understand and interpret objects, find their way and locate the information they require. This research seeks to establish the relation between certain factors in CMV information provision, and the navigation experiences of people. The initial results are promising in that they appear it show a consistent improvement by implementation of design heuristics. A more rigorous experiment will permit evaluation of a CMV under controlled conditions, in order to identify how users experience difficulty in navigating information spaces and why this happens.

### REFERENCES


