Inclusive Data Visualisation: A Multidisciplinary Approach

Hua Dong[1], Gordon Barr[2], Elizabeth Blackburn[3], Melissa Grant[4], Paul Piwek[5], Paul Shepherd[6], Nick Collins[7]

1. *Inclusive Design Research Group, School of Engineering and Design, Brunel University, Uxbridge UB8 3PH, UK*
2. *Department of Chemistry, University of Glasgow, Glasgow, G12 8QQ, UK*
3. *School of Physics and Astronomy, University of Birmingham, Birmingham, B15 2TT, UK*
4. *Dental School, University of Birmingham, Birmingham, B15 2TT, UK*
5. *Center for Research in Computing, the Open University, Milton Keynes MK7 6BJ, UK*
6. *Department of Architecture and Civil Engineering, University of Bath, Bath, BA2 7AY, UK*
7. *School of Informatics, University of Sussex, Brighton BN1 9RH, UK*

Abstract
Data are important sources of information and knowledge. To explore a more inclusive means of communicating data, a team composed of seven researchers in the UK from different disciplines conducted a series of workshops: the first to share state-of-the-art data visualisation techniques in various disciplines and to identify data visualisation challenges; the second to extract universal principles from good examples of data communication, to identify data visualisation criteria, and to develop a set of strategies for inclusive data communication. On the evidence of the first two workshops, we believe there is currently a great lack of inclusivity in data visualisation. Communicating data using multiple modalities (visual, auditory, haptic…) and understanding users' needs and expectations were proposed as the most important strategies for making data communication more inclusive for different target users. The third workshop will focus on developing and evaluating a methodology for more inclusive data communication with different groups of users (including people with disabilities).

Keywords
Data visualisation, inclusion, multidisciplinary collaboration

INTRODUCTION
Data are important sources of information and knowledge, and they take different forms, e.g. numbers, graphs, images, or texts. Data visualisation is the process of transforming data into sensory stimuli, usually visual images (Schroeder et al, 2003). The "main goal of data visualisation is to communicate information clearly and effectively through graphical means." (Friendly, 2008). Through effective visualisation, data can be rapidly understood by the user.

Good data graphs, as suggested by Edward Tufte (1992):

- Help the audience think about the important messages from the data, rather than about methodology, or something else;

- Avoid distorting what the data have to say;
- Present many numbers in a small space – but also emphasise the important numbers;
- Make large data sets coherent, and encourage the audience to compare different pieces of data;
- Reveal the data at several levels of detail, from a broad overview to the fine structure.

Effective data visualisation does not only facilitate learning, but also enriches the process of scientific discovery and fosters profound (and sometimes unexpected) insights. A good example is the periodic table of the chemical elements. To extract new meaning from the sea of data, scientists have begun to embrace the tools of visualisation (Frankel and Reid, 2008). The Cambridge Engineering Selector (CES) for material selection is an excellent example that illustrates Tufte's aforementioned characteristics of good data graphs. By visualising hundreds and thousands of materials in a novel way, it opens a new path to understanding materials and a new approach to material selection (Ashby, 2005).

To help data originators communicate data more effectively, Harvard University has recently organised a series of workshops on Image and Meaning (http://www.imageandmeaning.org/), aiming to "help scientists, writers and visual communicators develop and share improved methods of communicating scientific concepts and technical information through images and visual representations." The power to visualise and graphically represent results and ideas in multiple dimensions and to manipulate data has already been predicted as the next big revolution in technology.

To explore a more inclusive means of communicating data, seven UK university-based researchers started a multidisciplinary research collaboration project in May 2009. Their backgrounds include industrial design and engineering, architecture and civil engineering, chemistry, physics, biochemistry, computational linguistics and digital music. They all deal with a variety of data on a daily basis, such as numbers, codes, symbols, spectra, text, diagrams, tables, images, sound, and animation, and they all shared an interest in exploring effective methods of distilling meaning from data.

The collaborative project has provided an opportunity for the researchers to apply inclusive design thinking across different disciplines, and to evaluate how 'useful', 'usable' and 'inclusive' data from any experiment can be made. The project aimed to:

1) identify data visualisation challenges in different disciplines and data communication criteria from the perspectives of both data developers and data users
2) propose strategies in making data communication more inclusive for different target users
3) develop a methodology for evaluating the effectiveness of data communication.

The multidisciplinary dialogue about data communication was conducted through a series of three workshops over a period of 18 months (the first two held in August 2009 and February 2010; the third workshop to be held in November 2010), involving the seven researchers, invited guests (e.g. from areas of Brain Scans and MRI, Archaeology, Architecture, Semantic Web, and Software Design), and different types of data users (e.g. knowledgeable professionals, layman users including people with disabilities).

DATA VISULISATION CHALLENGES

At the first workshop (August 2009), each researcher presented a short talk in a Pecha Kucha style (20 slides with 20 seconds per slide) on what constitutes 'data' in each of their own disciplines. They also gave examples of the types of data and discipline-specific data communication methods and techniques. For example, in chemistry, data are often presented in 1D, 2D or 3D forms; while in architecture, they are typically displayed in 2D (layout), 3D (models) and 4D (animation). Figure 1 shows some examples of data presented at the workshop.

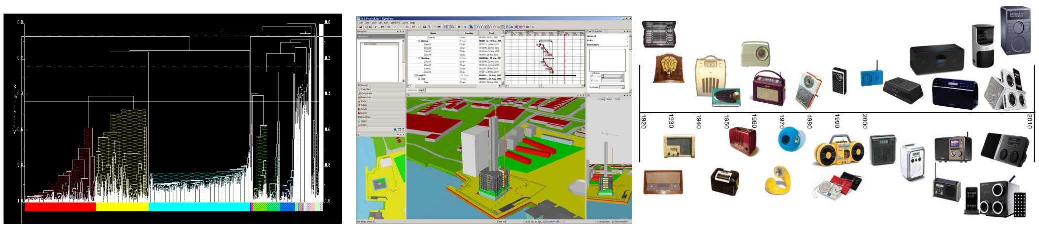

Figure 1. data examples in chemistry, architecture and design

A number of data visualisation challenges were identified, including:

- Communicating data to different potential users (e.g. expert colleagues; knowledgeable professionals, layman clients and interested parties)
- Difficulties in choosing the right (or optimal) tool for complicated data sets
- Balancing clarity against economy (e.g. the amount of texts used)
- Different terminologies
- Dilemmas in applying novel data visualisation methodologies (graphic design, the technology of graphic production etc): they may help attract users' attention but may also distract users from focusing on important messages from the data
- A huge increase in the amount of data available because of technological facilitation, yet which corresponds in many cases only to electronic versions of standard practice.

DATA VISUALISATION PROCESSES

The first workshop also saw the researchers' engagement with a data visualisation exercise: an 'unknown' large dataset (numbers in a table format, with no context apart from names of countries and continents and the time period) had been selected from a publically available data source, and each researcher was asked to analyse the dataset prior to the workshop, visualise the data and present a poster outlining the approach they had taken to analysing and presenting this unknown dataset. As a result, seven different versions of 'visualisation' of the same dataset were demonstrated at the workshop, ranging from typical EXCEL bar charts to sophisticated cluster analysis graphs (dendrogram), and the original data (i.e. numbers) were also transformed into different modalities: 3D models, sound, and spoken language.

Follow-up analyses of the processes of visualisation were conducted after the first workshop, and the pros and cons of each method were discussed at the second workshop. Figure 2 shows the visualisation posters (upper row) and the comments on pros and cons (lower row). The pros were written on green sticky notes, and the cons in orange.

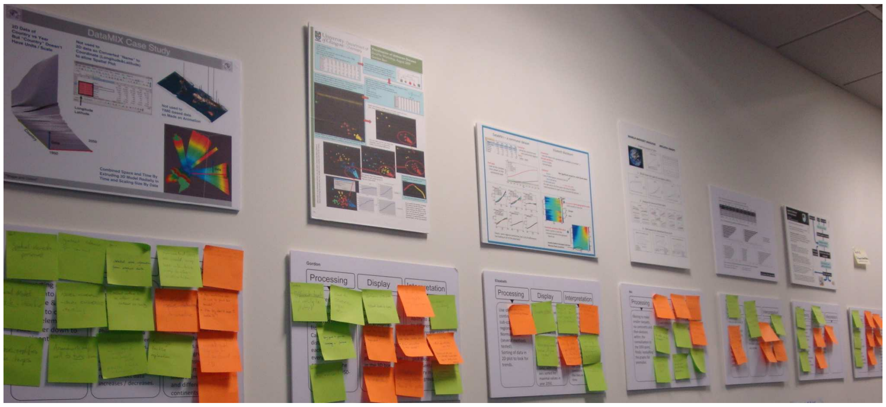

Figure 2. Data visualisation posters and the comments on pros and cons of the methods used in each poster.

Although the researchers from different disciplines used different techniques in visualising the given dataset, common patterns were identified (see Figure 3). The common procedure was to process data based on the source (raw data), through deducting noise ("remove/select/filter/extract/distil"), sorting ("sort/cluster/remap/group/organise") and normalising ("calculate/normalise/convert"). Sometimes new data were also created in the process. When visualising data, common methods adopted included 2D graphs, 3D plots, change of modalities (e.g. visual/auditory/dialogue), and adding new dimensions (e.g. colour, animation)

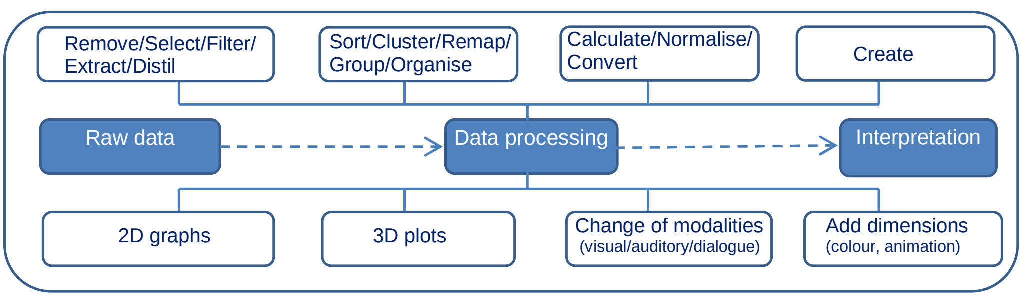

Figure 3. a common pattern of the data visualisation process

DATA VISULISATION EXAMPLES

In the second workshop, each researcher presented good and bad data visualisation examples from their own field, and Figure 4 shows an example from biochemistry. On the left is an example of good communication in biochemistry: colour and spatially divided items with distinct summaries of information; while the right hand side shows a bad communication in biochemistry: mono chrome and spatially confused items with indistinct actions.

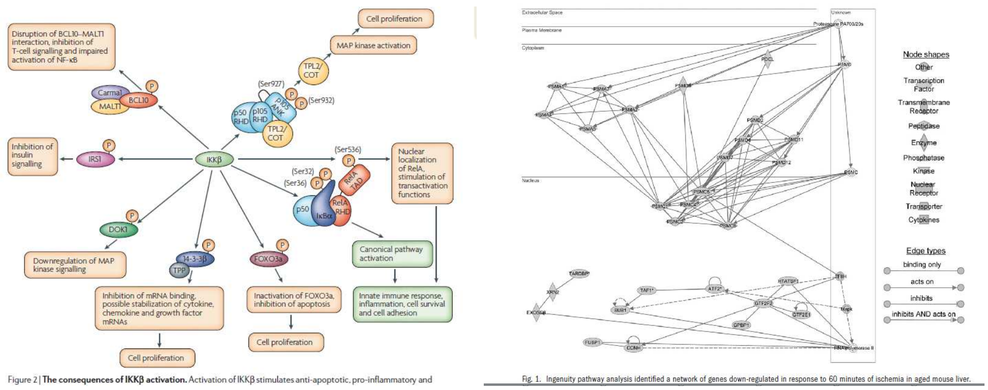

Figure 4. Good and bad data visualisation examples in biochemistry

A creative exhibition of live, general examples of data visualisation was also organised at the second workshop (Figure 5).

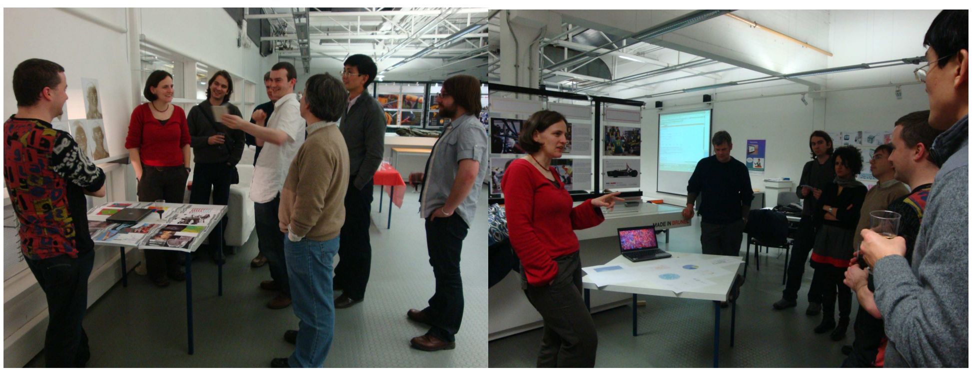

Figure 5. Creative exhibition of data visualisation examples

Based on the examples, a number of characteristics of good data visualisation were identified:

- Clean presentation with clear labels and use of standard conventions
- Showing contexts and details-on-demand
- Allowing inspection by giving access to the underlying data
- Appropriate scale with size of data set
- Concise explanation (annotation) with distinct summaries of information
- Highlighting important data
- Adding dimensions (e.g. colour) to make data patterns clearer
- Interactive features to enable multiple views and data references
- Showing right level of information.

DATA VISULISATION CRITERIA

An important aim of the research was to identify data communication criteria from different perspectives (e.g. expert colleagues; knowledgeable professionals, and lay users). In the second workshop, the participants were asked to select up to 15 most important data communication criteria from a list compiled based on the issues mentioned at the first workshop. Figure 6 uses a 'tag cloud' tool (available from www.manyeyes.com) to illustrate the criteria in terms of their relevant importance.

The relative size and weight of the font for each criterion corresponds to the relative frequency of its mention by the participants.

Figure 6. Data visualisation criteria: researchers' viewpoint

A short questionnaire was taken to the 5th Cambridge Workshop on Universal Design and Assistive Technologies (CWUAAT'10) in March 2010 where data users' viewpoints on data visualisation criteria were collected. The CWUAAT workshop was selected as it had an audience who were aware of inclusive design. The corresponding author gave a short introduction to the research project at the CWUAAT workshop's user forum session, and then asked the audience to write down their five most important criteria for data visualisation and three strategies for making data communication more inclusive. The results (based on 19 responses) of the criteria are shown in Figure 7.

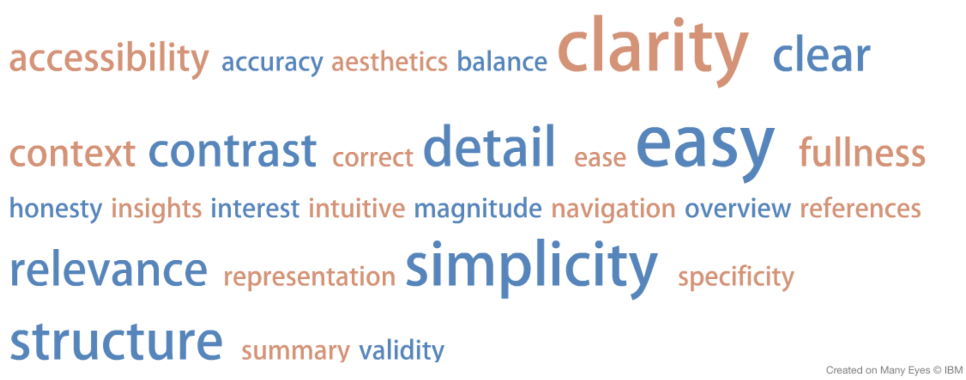

Figure 7. Data visualisation criteria: knowledgeable professionals' viewpoints

Although intended for lay users' viewpoints, the backgrounds of the questionnaire respondents suggested that they were more "knowledgeable professionals" than layman users. Among the 19 respondents, six were "data developers/researchers", three were "data users", and nine were "both data developers and data users".

It is planned that lay users' viewpoints on data visualisation criteria will be collected from the third workshop in November 2010.

DATA VISUALISATION STRATEGIES

In the second workshop, the participants were also each asked to propose three strategies for making data visualisation more inclusive to different groups of users. The results (based on ten responses: the seven authors and three workshop speakers) are shown in Table 1. The CWUAAT participants were also asked to do the same task, and the results (based on 19 responses) were summarised in Table 2.

Table 1. Strategies proposed by researchers

Strategies	**Numbers of mentions**
multiple modalities, views, media	7
multiple levels of detail, complexity, ability to drill down	4
involve user	3
interactive	3
use of metadata/standards	3
multilingual labelling	1
use the ‘real thing’, rather than data	1
protect privacy	1
allow users to tailor visualisation	1
portable tools for multimodality	1
explain purpose clearly	1
selecting appropriate tools	1
understand user diversity	1
least ‘capability demand’	1
extreme users	1
accuracy	1

Table 2. Strategies proposed by knowledgeable professionals

Strategies	**Numbers of mentions**
understand users	6
avoid data overload/keep it simple	5
multi-modal presentation/flexibility of visualisation	4
fond size must be adequate	4
define objective and customise for purpose	4
user involvement	4
highlight salient points/emphasise important information	3
accessibility /utilise HCDI guidelines	3
inclusion of raw data/data source	2
good contrast of text and background	2
build up understanding /interpret what is means	2
different levels of detail	2
be more graphically illustrative	1
use simple diagram in support of text	1
the structure of the content presented is important	1
3D	1
Interactive	1
consider context	1
possibility to navigate through contents (pick and mix)	1
reliability	1
use clear and simple languages (in text)	1
clear labelling of information	1

DATA VISULISATION METHODOLOGIES

The participants formed three groups during the second workshop to discuss the methodologies of data visualisation. Figure 8 shows the results presented by two groups. The ‘PURRFECT PROCESS’ on the left suggests that data interpretation has to be connected to both the data and the user, and the efficiency of the communication has to be evaluated by the user. The methodology on the right put data in a research project context, suggesting that data and research questions are interrelated. Data visualisation is for internal use (to help the researcher better understand the data), and data communication (for external use, i.e., to communicate research to all parties interested) and should take all potential users into consideration. The links between data communication and the original data have to be strengthened, and user feedback should be used to improve data visualisation, and consequently, data communication.

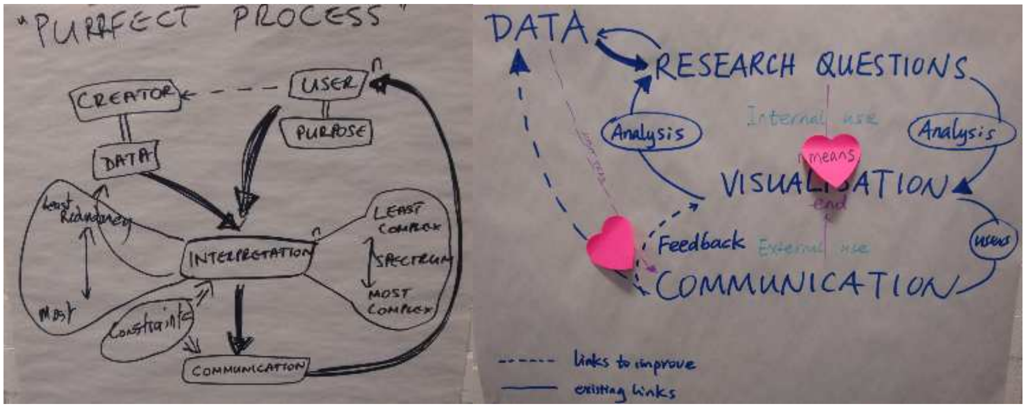

Figure 8. Data visualisation methodologies

Figure 9 shows the third group’s proposed methodology framework.

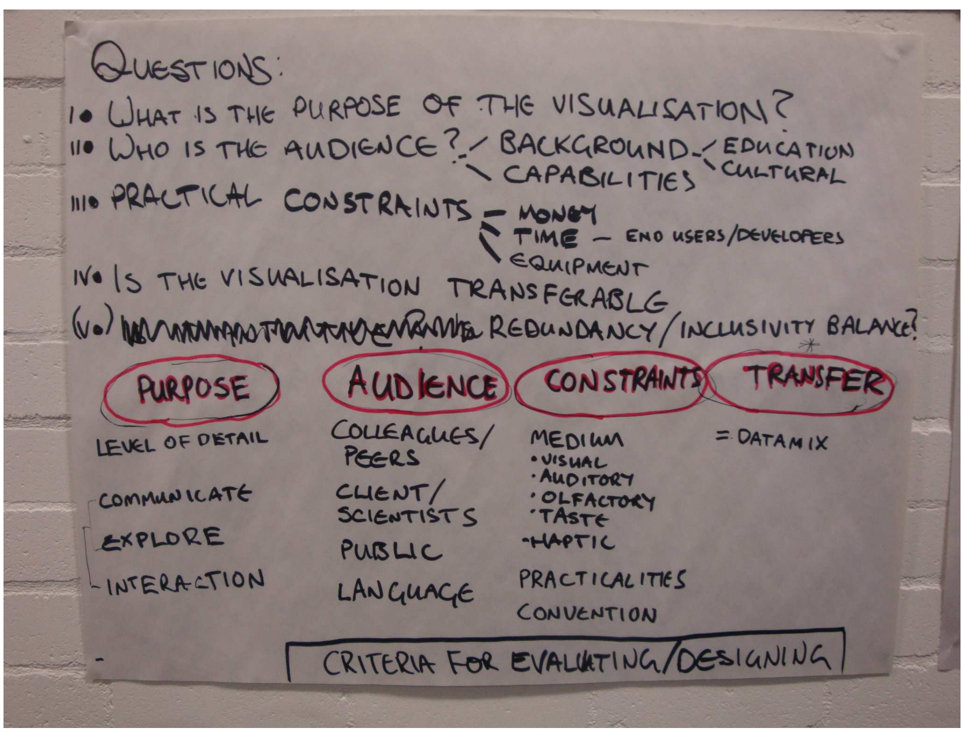

Figure 9. Data visualisation methodology framework

The framework started with a set of questions:

What is the purpose of the visualisation?

Who is the audience? (background: education; cultural; capabilities)

Practical constraints (money, time: end users/developers, equipment)
Is the visualisation transferable?
Redundancy/inclusivity balance?

Therefore, to evaluate data visualisation, the following aspects should be considered:
Purpose (this will help define the level of detail to be communicated)
Audience (e.g. colleagues/peers, client/scientists, public, use of language)
Constraints (medium: visual, auditory, olfactory, taste, haptic; practicalities, convention)
The group believed that this research project would generate insights that will help with the transfer from raw data to effective communication.

DISCUSSION

This multidisciplinary project has helped identify a number of data visualisation challenges. Our primary point is that there is currently a great lack of inclusivity in data visualisation. Even within narrow expert communities (e.g. biochemistry), data visualisation can be exclusive/difficult to understand without expending significant effort.

The researchers and the knowledgeable professionals share some common strategies for making data visualisation more inclusive to different groups of users, for example, “multi-modal presentation”; “involving users” and “use of accessibility standards/guidelines”. The researchers frequently mentioned the strategy of providing “multiple levels of detail” and making the data visualisation “interactive”; while the knowledgeable professionals think “understanding users” is of paramount importance; and they frequently mentioned the strategy of avoiding “data overload” and keeping the information “simple”.

This was in interesting contrast to the comments by the researchers which mentioned ‘complexity’ as a point of interest, expressing the desire to communicate complex information successfully. We suggest that, here, ‘complexity’ and ‘simplicity’ represent the same ideal: complex interpretation, rendered accessible for the end user.

The most important data visualisation criteria, as suggested by the researchers, were concerned with “clarity”, “(level of) complexity”, “context”, “explanation”, “multiple (views)”, “purpose”, “relevance” and “trust”; while the knowledgeable professionals at the CWUAAT workshop suggested that the most important criteria included “clarity”, “detail”, “ease of use (easy to understand/remember/identify/compare)” “simplicity” and “structure”.

Although people from different disciplines use different techniques in visualising data, common patterns and procedures (as shown in Figure 3) were identified. One common theme was the addition of extra dimensions to assist understanding. This can be done through a variety of means:

- Physical objects
- Time
- Colour
- Sound

All of these possibilities have advantages and disadvantages. For example, colour can help make data patterns clearer, and provides an easy way to delineate between differing data streams. However, it may introduce unintended exclusion, e.g. for the colour blind, or for those without access to colour printers. In some cases, this can be resolved by application of alternative markers (e.g. spatial separation).

The researchers participating in the project all emphasised the importance of having access to raw data; this may be because many of them not only use data but also generate new data (and visualisation) based on combining datasets.

The data visualisation methodology will be further developed and evaluated with the potential data users at the third workshop in November 2010.

CONCLUSIONS

There is a lack of inclusivity in data visualisation, but not much research on this topic. This research project has taken a multidisciplinary approach to explore data visualisation from the 'inclusion' viewpoint. "Clarity" has been identified as the most important, and commonly agreed, criterion for data visualisation. While data developers think it is very important to provide "different levels of complexity", "multi-views" and the "context" of data; data users are more concerned with the "ease of use", "simplicity", "detail" and "structure". Communicating data using multiple modalities (visual, auditory, haptic…) and understanding users were regarded as most important strategies for making data communication more inclusive for different target users.

REFERENCES

Ashby, M. (2005) *Material selection in mechanical design* (3rd Edition), Butterworth-Heinemann

Frankel, F. and Reid, R. (2008) "Big data: distilling meaning from data", *Nature*, 455, 30

Friendly, M. & Denis, D.J. (2001). "Milestones in the history of thematic cartography, statistical graphics, and data visualization". Web document, http://www.math.yorku.ca/SCS/Gallery/milestone/. [Accessed on the 30th May 2010].

Schroeder, W., Martin, H., and Lorensen, B. (2003), *The visualisation toolkit* (3rd Edition), Prentice Hall PTR, New Jersey

Tufte, E. (1992), *The visual display of quantitative information*, Graphic Press

ACKNOWLEDGEMENTS

This research project has received support from the UK's National Endowment for Science, Technology and Arts (NESTA), an independent body with a mission to make the UK more innovative. The authors would also like to thank the following people for their input to the project: Professor Ann Heylighen from K.U.Leuven, Belgium; Professor Yong Zhang from Tsinghua University, China; Dr Michael Wermelinger from the Open University, UK; Ms Farnaz Nickpour, Mr Chris McGinley and Mr Abdusselam Cifter from Brunel University, UK; and the CWUAAT workshop participants who took time to answer the questionnaires.