The KINDS Project

Keith Cole[1], *Chun S. Li*[1], *Adrian Moss*[2], *James Petch*[2], *Yau J. Yip*[3], *Andrew Johnston*[3], *Kamie Kitmitto*[1] and *Andrew Basden*[3]

ABSTRACT: *The KINDS project is being funded under the JISC New Technologies Initiative to develop a Knowledge based Interface to the National spatial Data Sets (KINDS), such as Bartholomew digital map data, available on-line via MIDAS (Manchester Information Datasets and Associated Services). The aim of the project is to increase the use of national spatial datasets by educating existing and potential users in how to use them in teaching and research as well as improving accessibility. An experimental WWW server (http://midas.ac.uk/kinds/) has been established which allows users to perform spatial queries using search engines and clickable maps as well as providing an interface for authorised users wishing to undertake real time processing of data using the Arc/Info GIS running on the Cray Superserver CS6400.*

1 Introduction

Manchester Information Data sets and Associated Services (MIDAS) is based at Manchester Computing, Manchester University and provides a range of national dataset services to the UK academic community. The service provides flexible on-line access to a wide range of large and complex teaching and research datasets, such as the UK 1991 Census of population statistics, Bartholomew digital map data, Landsat and SPOT satellite images. MIDAS also provides a range of specialist support services on these datasets to a large and diverse group of users with varying levels of awareness and/or technical expertise. MIDAS, which runs on a Cray Superserver CS6400, is jointly funded by the Joint Information Systems Committee (JISC) of the Higher Education Funding Councils, the Economic and Social Research Council (ESRC) and the University of Manchester and is freely available for academic use throughout the UK. To date, over 2500 researchers from 112 institutions are registered to use the service.

Despite the provision of flexible on-line access to a range of large and complex spatial and spatially referenced datasets and associated support services the uptake of these datasets in teaching and/or research has been found to be limited (MIDAS Annual Report, 1995). Accessibility and usability of spatial data sets are major bottlenecks to increasing the number of applications (Petch *et al*, 1995; Li *et al*, 1995). Users must be able to identify and locate suitable data and be educated in the appropriate use in order to exploit available data resources effectively.

To increase the accessibility of data, data providers must promote awareness of the existence and contents of those data sets. Many spatial data sets are of potential use to a range of different end users however if potential users are unaware of the existence of a data set the result may be low utilisation and even a waste of an expensive resource (Cornelius and Strachan, 1989; Ruggles 1990; Walker *et al*, 1992). The provision of effective meta data (data about data) underpins attempts to increase data accessibility (McLauglin and Nichols, 1994). The development and deployment of meta information systems is currently receiving global research attention. See for example GENIE (Walker *et al*, 1992), GeoWeb (Plewe, 1994) and Project Alexandria (Frew *et al*, 1995; Andresen *et al*, 1995).

Poor usability results from low understanding of the use of spatial data. Spatial data analysis is a knowledge intensive activity. New users face a significant learning curve when adopting spatial data. The techniques associated are often significantly different from other types of analysis (Fotheringham and Rogerson, 1993). Users must familiarise themselves with the command structure and nomenclature of geographical information systems. Often users must also have expertise in data pre-processing and the systems of data provision used by data providers. Often data sets have to be used together with ancillary software packages to support operations such as data handling, map plotting, etc. To be able to use the

[1] Manchester Computing, University of Manchester, Manchester, UK
[2] Department of Environmental and Geographical Sciences, Manchester Metropolitan University, Manchester, UK
[3] ITI Institute, University of Salford, Salford, UK

data, the user should have both knowledge of data coverage, structure and the support systems. The combinations of high level technical skills required often result in spatial data handling being the preserve of the highly technically competent.

The Knowledge-based Interface to National Data Sets (KINDS) project aims to extend and intensify the use of national spatial datasets available from MIDAS by educating potential users in their use. This is a dual mission. Firstly potential users are to be made aware of the existence of the data sets and secondly guided in their use. In order to achieve this, KINDS must reach the widest possible audience through the World Wide Web (WWW) with user friendly, easy to access and effective Internet-based search engines and secure hypertext interfaces for user to browse and handle spatial data.

This paper presents some selected findings from two surveys of potential users examining actual and potential use of national spatial datasets; users' knowledge of computing and information systems and human-computer approaches to spatial data. The survey findings have been used as the basis for the development of KINDS experimental system. The paper describes the design and development of a series of interfaces on WWW to enable the user to easily discover information about spatial data services and example applications, and carry out real-time task processing with ESRI Arc/Info running on the CS6400 to generate maps over the Internet.

2 KINDS technical survey and findings

An extensive user survey including both semi-structured interviews and a technical questionnaire survey was carried out in Manchester at the early stage of the KINDS project to examine the academic requirements for data in teaching and research, and approaches likely to be used to access spatial data.

The Manchester academic community is amongst the largest ones in Europe. Seventy eight researchers, lecturers and post-graduate students at academic departments representing about 28 disciplines were approached to determine their needs for spatial data. The rationale of the user survey was to reveal the extent of data use within the academic community and paths of dissemination through users. Hence a mixed approach of quantitative and qualitative research methods was adopted. One of the prime objectives of the survey was to examine informal methods that users employed to gain support. This is the phenomena of a user who when faced with a problem using a computer system opts to speak to a nearby colleague or friend rather than approaching formal support services. In order to test the extent to which this took place in the data processing environment respondents were asked to suggest likely colleges to be interviewed. Whilst we fully acknowledge that this process does not adhere to random sampling rules and precludes rigorous statistical analysis a fuller picture of information dissemination within higher education was revealed as a result. The full results of the user survey will be reported in a future publication. We present some of our analysis here as background to what will follow.

2.1 Actual and potential use of the data

Use of spatial data is a knowledge intensive activity. Spatial analysis concepts and techniques are significantly different from those used in other analysis methods (Fotheringham and Rogerson, 1993). Users must familiarise themselves with the command structure and nomenclature of geographical information systems, such as Arc/Info, and the environment in which they reside - often GIS are mounted on high end UNIX workstations whilst the majority of new users are familiar with personal computers.

The survey revealed that there was a considerable mismatch between actual and potential use of data (see figure 1). Firstly, over 50% of respondents were not using any of the datasets available via MIDAS. Of those respondents who did use data in their teaching and/or research (see the left hand columns in Figure 1) the majority were using either the Census or the major government and other continuous surveys. It was evident that very little use was made of any of the national spatial datasets

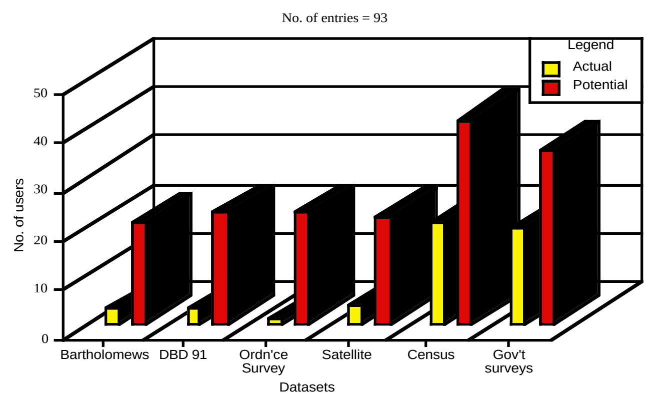

Figure 1: Actual use vs potential use of data

available to the UK academic community, such as Bartholomew digital map data or satellite data. However the picture changes when respondents were asked about their potential use of data (see the right hand columns in figure 1). As soon as they became aware of the content, coverage and availability of the various national datasets most respondents felt that they would be a useful resource for teaching and/or research. However the most significant potential increase in use was for national spatial datasets.

2.2 The user's technical ability

Knowledge of information technology seems to be a key factor which decides whether data set applications are successful. To investigate how these users access data and how user friendly and effective interfaces are for spatial data handling, a technical questionnaire survey was carried out. User knowledge, interest in networked information systems and human computer interaction issues were examined. Forty one (52.5%) respondents returned completed questionnaires. The users were classified into four categories (basic, intermediate, experienced and professional) based on knowledge and daily use of information systems and computer-based applications.

The extensive use of the Email facility indicated that most users already have access to computer networks. The Internet utilities (telnet, FTP, etc) were also widely used. WWW, the most recent Internet service, has become one of major platforms in the academic community. Over 70% respondents commented that WWW has been a useful resource for their teaching and research. For each category of user the following table shows usage of the Internet-based Information systems.

Table 1. The Internet-based information systems being used

LEVEL - user level by NETINFO - usage of networked information systems

LEVEL	Count	news group 1	gopher, wais 2	WWW 4	WWW + others 6	Row Total
basic	1		1	3	2	6 17.1%
intermediate	2	5	1	1	10	17 48.6%
experienced	3				10	10 28.6%
professional	4				2	2 5.7%
Column Total		5 14.3%	2 5.7%	4 11.4%	24 68.6%	35 100%

Number of Missing Observations: 6

In terms of the potential use of the MIDAS service it was significant that the vast majority of respondents used PCs with MS Windows™ and made little use of UNIX based systems such as the CS6400.

2.3 User interfaces

The user interface is one of the most important aspects of a spatial information system. Respondents were asked to select the types of interfaces they preferred. Selections were listed in sequence according to the respondents' preference. The results were processed using weighting and ranking techniques, to assess which class of users found which interfaces more useful. Table 2 lists the interfaces ranked by the points scored.

Some users were unable to comment on interfaces since they had no experience. To address this data on the respondents' background was considered. The following table shows the ranking indicated by the average points scored.

2.4 Bottlenecks affecting use of spatial data

The survey revealed that many respondents were unaware of the availability of national spatial data sets on-line from MIDAS. Factors such as the significant spatial data handling learning curve, inadequate technical guidance and support help to explain the lack of use. In addition, the lack of UNIX expertise should not be underestimated as a significant deterrent to use for a large proportion of potential users.

Ease of use of user interfaces was preferred over functionality by users with less technical skills. The results of the survey were broadly in compliance with Davis and Medyckyj-Scott (1994) who reported that inexperienced users suffered difficulty in transferring existing knowledge from their disciplines to spatial data handling. The results indicate that user friendly and flexible interfaces are important for improving spatial data accessibility and usability.

3 Development of interfaces for browsing spatial datasets

3.1 Bartholomew Map Dataset

The Bartholomew digital map data has been used as the test dataset by the KINDS project. It is a layered vector map data set comprising of point, line and area features. The data is structured into several sets comprising World, European, GB, London and Central London coverages. In the Bartholomew (Great Britain) data set, the coverage is divided into tiles, based on the National Grid. Each tile, covering an area of 100 km square, is identified by a set of two characters as shown in Figure 3.

The data for either the GB national coverage or a tile is stored in 16 thematic data layers including administrative boundaries, contours, roads, railways and ferries, point features, urban areas, and water, etc. In addition, there is an annotations layer that contains textual information describing the feature data.

Within the Bartholomew data set, the features are organised into classes, each class is identified by an OBS_ACC_NO (observation accession number), which describes the feature and its entity types (point, line or polygon). The OBS_ACC_NOs uniquely identify each type of feature. For instance in the 'Roads'

Table 2. Answers ranked by the total points scored

Type of Interfaces	Std Dev	Minimum	Maximum	Sum	Valid No
Interactive Dialog Boxes	1.24	2	6	184.00	36
Tree Structured Menus	1.19	2	6	124.00	29
Interactive Maps	1.28	2	6	121.00	27
Multi-Selectable Items	1.15	3	6	102.00	24
Natural Language Interfaces	1.35	2	5	71.00	23

Table 3. Answers ranked by the average points

Type of Interfaces	Mean	Std Dev	Minimum	Maximum	Valid No
Interactive Dialog Boxes	5.11	1.24	2	6	36
Interactive Maps	4.48	1.28	2	6	27
Tree Structured Menus	4.28	1.19	2	6	29
Multi-Selectable Items	4.25	1.15	3	6	24
Natural Language Interfaces	3.09	1.35	2	5	23

data layer, existing motorways are the unique code no (235), the primary trunk dual carriage way A roads are referenced as (173291) and so forth.

The following is a list of thematic data layers available in the Bartholomew (GB) dataset,

Administrative Boundaries (National and regional/county boundaries, lochs (lakes) and the coastline)

Contours (contours at 100m intervals plus 50m and 150m)

Danger Zones

Drainage (including canals)

Forest Parks

National Parks

National Trusts

Other Lines

Points (Leisure, physical, road, industrial, other transport and road distance points, etc.)

Rail and Ferry Links

Roads (Motor ways, A, B roads, minor roads and roads under construction)

Regional Parks

Scenic Areas

Topography (Rocky shores, beaches and woodland)

Urban Areas

Water (lochs (lakes) and marshes)

Annotation (Cartographically placed text annotation).

3.2 The KINDS thematic map library

Potential spatial data users can gain information about the data set far more easily by browsing its contents than by reading a textual description. Thus enabling users to quickly browse through maps was a major objective in increasing awareness of the Bartholomew map data. The WWW is a fast and feasible way of presenting maps by distributing GIF format images. Such images which are of sufficiently good quality to display spatial objects but require only small amounts of memory and transmission time through the Internet to the user's client software.

A number of sample maps were generated manually using ESRI Arc/Info Arcplot in the early stages of the KINDS project as a feasibility study. After the KINDS experimental system was released on WWW, users expressed an interest in seeing more detailed feature maps covering more specific areas. A feature map library of the major features contained in the Bartholomew (GB) map data was built to provide as detailed spatial information as possible. Spatial features are linked to their corresponding textual descriptions in hyper text markup language (HTML) pages. The features and their descriptions can be retrieved by either an interactive map interface or a search engine.

A virtual map library has been created. The library comprises of a full UK national coverage directory and 55 (for all 100km squares) sub-directories named after the corresponding tile of the National Grid. ESRI Arc/Info Arc Macro Language (AML) was used to create scripts to automatically generate feature maps from the data set and generate legends automatically. Arcplot is unable to export maps in GIF format and so SDSC Image Tools (produced by San Diego Supercomputing Center) was used for converting maps from Sun Raster to GIF format.

3.3 Map interface

An interactive map (also sometimes referred to as a clickable image) is an inline image in an HTML document. The position of any mouse click within the image is captured using of the HTML tag ISMAP. When a user clicks the mouse over the image, the browser sends the pixel co-ordinates to the WWW server. The co-ordinate information is then processed by a program on the server to return an appropriate URL (HTML document) after comparing the mouse co-ordinates with boundary location information in the imagemap database. The ISMAP tag provides for a limited degree of spatial querying of maps.

The required number of co-ordinates is dependant upon the shape of region to be defined. For circles, two pairs of co-ordinates are required: centre and any edgepoint; for rectangles, co-ordinates of upper-left and lower-right; for polygons of 100 vertices at most, each co-ordinate pair stands for a vertex.

The ISMAP facility is a powerful tool for producing WYSIWYG - "what you see is what you get" interfaces. The interactive map provides an intuitive and easy to use method for the user to swiftly browse through the Bartholomew data's "layer-and- tile" structure. The interactive map is based on a UK map using the National Grid, regional and county boundaries for geographical location referencing (Figure 3). Countries are marked in different colours for users to easily point to an exact area of interest, for example, Scotland is in blue colour, England in red and Wales in green. The user can simply move the mouse to an area (tile), and click on it to see detailed thematic data which is linked to the KINDS map library with over 800 feature maps.

3.4 Free text search engine

A "free text" search engine provides a direct entry for both expert and inexperienced users to quickly discover information about the data set. The interface uses a dialogue box for the user to enter a query which is passed to the search engine. The engine then retrieves an index file linking to the KINDS thematic map library and returns headings of documents which match the users query.

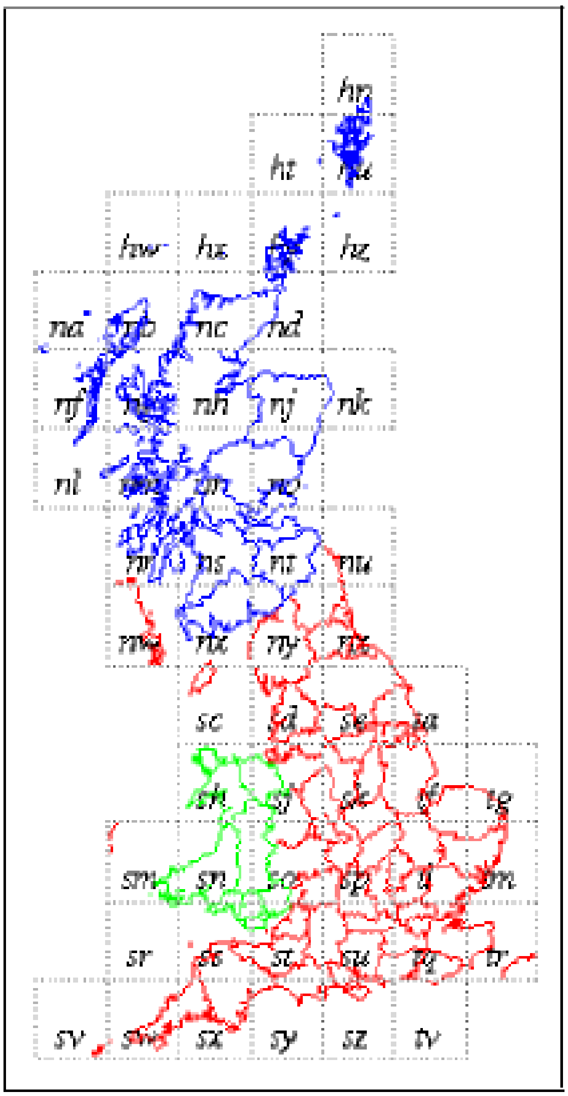

Figure 3: An interactive UK map. Copyright: Bartholomew

The index file is an important element in the free text search engine. Spatial features originally organised based on National Grid references have been restructured to be linked with UK counties/regions, and major cities. Each entry in the file is indexed with both the two-character identification of tile and counties or regions as well as major cities covered by the tile. Thus the search engine complements the map interface by allowing the user to search for information about specific cities rather than using the technical tile structure.

The search engine executes a Perl script in the server. It filters out irrational search terms (with less than two characters or starting with symbols such as "/", ".", "*") which may cause inaccurate results to be presented. Boolean (logic AND, OR) searches and right-hand truncation searches (with an "*" at the end of search term) are supported to make searches effective and efficient.

Help information including a list of indexed terms used in the search engine and information about its structure have been added for inexperienced users. The user is also able to start querying by clicking on 'hyperlinks' listed in the help file.

4 Making Road Maps - an example for directly handling spatial data over WWW

4.1 Representation of UK Roads

A dynamic link between the WWW and the Bartholomew data using ESRI Arc/Info has been created to familiarise the user with working with spatial data. The Bartholomew (GB) data set is structured into 16 thematic data layers comprising over a thousand classes of features, the roads data layer was selected for the dynamic link experiment. The roads layer includes 22 classes of road features (see the list below) and each feature is available in both the national coverage and individual tiles.

The UK roads are well classified and structured in the dataset. The following taxonomic conceptual model (Figure 4) shows the hierarchical relations of roads.

The Bartholomew data is mounted on the MIDAS national data sets machine as Arc/Info coverages. To assess the suitability of the data set, users should first manipulate it within Arc/Info. Therefore users must already have considerable knowledge of spatial data handling before being able to make a decision about the use of MIDAS data. New users are thus placed at a considerable disadvantage. The KINDS experimental system reverses this situation by allowing users to manipulate the data set without having to directly use a geographical information system. A visual approach based on a clickable UK map (as in Figure 5) is adopted for choosing the coverage. The user can select features they wish to see on maps using a simple forms interface.

4.2 WWW form-based interface - a front end to the map maker

WWW forms (or fill-out forms) are a computer equivalent of paper forms. When users 'submit' a WWW form the users responses are transmitted to a program on the HTTP server for processing.

OBS_ACC_NO	Description	Feature Type
235	Motorway	Line
245	Motorway Under Construction	Line
173287	Motorway Tunnel	Line
173291	A Road Primary Trunk Dual C/W	Line
173292	A Road Primary Trunk Single C/W	Line
173293	A Road Primary Trunk Passing Places	Line
233	A Road Primary Non-Trunk Dual C/W	Line
226	A Road Primary Non-Trunk Single C/W	Line
173294	A Road Prim. Non-Trunk Passing Places	Line
231	A Road Non-Primary Dual C/W	Line
227	A Road Non-Primary Single C/W	Line
222	A Road Non-Primary Passing Place	Line
243	A Road Dual C/W Under Cons (all)	Line
241	A Road Single C/W Under Cons (all)	Line
232	B Road Dual Carriage Way	Line
230	B Road Single Carriage Way	Line
229	B Road with Passing Places	Line
244	B Road Dual C/W Under Cons (all)	Line
242	B Road Single C/W Under Cons (all)	Line
73	Road Tunnel	Line
130	Minor Road	Line
173295	Private Road	Line

A form provides for complex interactions between the user and other software via programs residing in the HTTP cgi-bin (executable programs) directory.

A forms interface to the ESRI Arc/Info system has been created to allow WWW users to interact with spatial data (Figure 5). The interface enables the user to create simple maps which inform about the contents and potential of the Bartholomew data set. The test application allows the user to select and display elements of the road data layer of the Bartholomew data set.

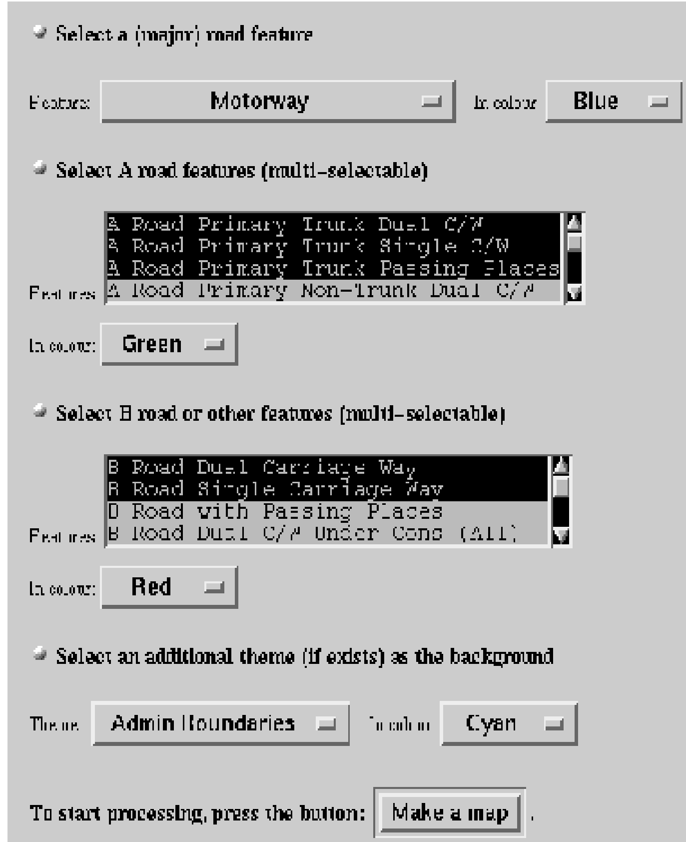

Figure 5: The forms interface

4.3 AML scripts coding - an automatic process

Submitting the WWW form activates Arc/Info after generating an AML script via a translation program residing in the HTTP servers' common gateway interface (CGI). The translation program is implemented using Perl. The query string is split into items to determine where to access the data (i.e. which tile);

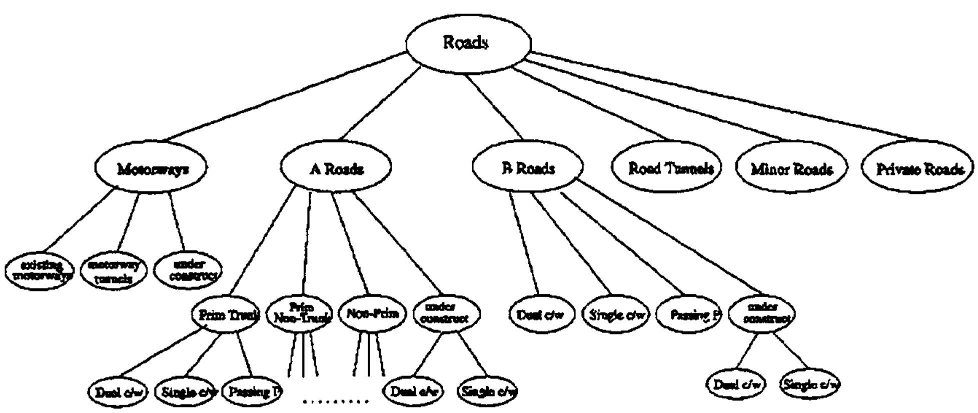

Figure 4: The hierarchical structure of roads

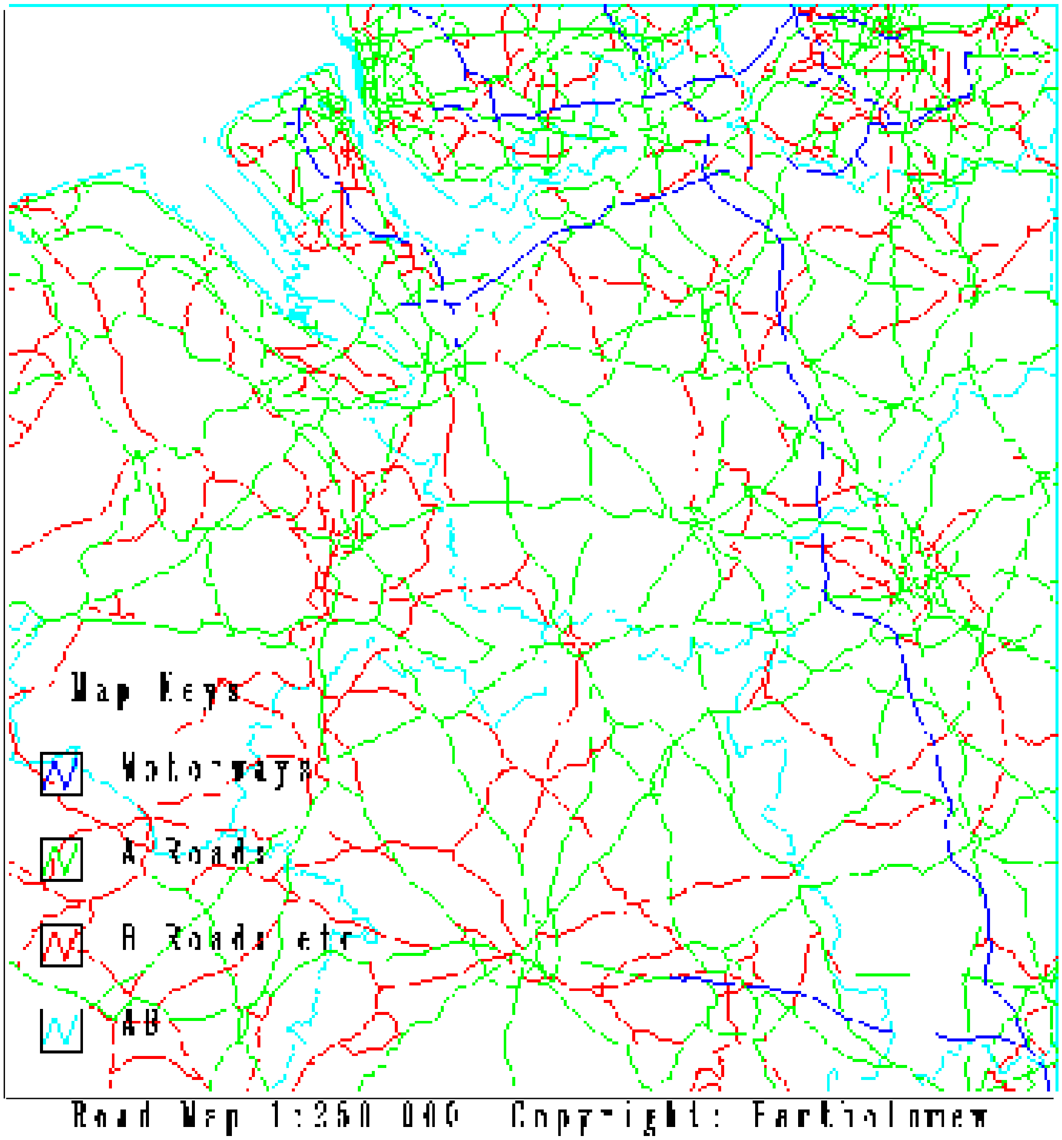

Figure 6: An example road map.

what road features and additional themes are selected and what colours should be used to mark these features.

Subsequently an AML script then can be coded to produce a map reflecting the user's request. A file comprising map symbols and textual descriptions for generating map keys (legends) with the AML script is also produced. The result of running the AML script is a 7" by 9" road map in encapsulated post script (EPS) format. The EPS file is then converted into GIF format by using image processing tools. The process in its entirety normally takes about 60-80 seconds to complete and send a map (9k to 21k in size) back to the user's WWW browser (this is dependant upon on the speed of the network). Figure 6 is an example road map about an area of south Manchester and Cheshire generated according to the user's request in Figure 5.

4.4 Outline of the KINDS Map Maker

The following schema illustrates the major processes of the KINDS experimental system, where the server-side executable Perl scripts set up the links between the HTTP server, Arc/Info, map data and other external packages running on the CS6400.

5 Future development

The KINDS experimental system has demonstrated the technical feasibility of building Internet-based interface systems to the MIDAS service as means of improving accessibility and usability of large and complex spatial datasets.

KINDS is a three year initiative at the end of the first 18 months of funding. In the remainder of the funding period further data sets and software packages will be added to the existing framework. The existing interface will be comple-

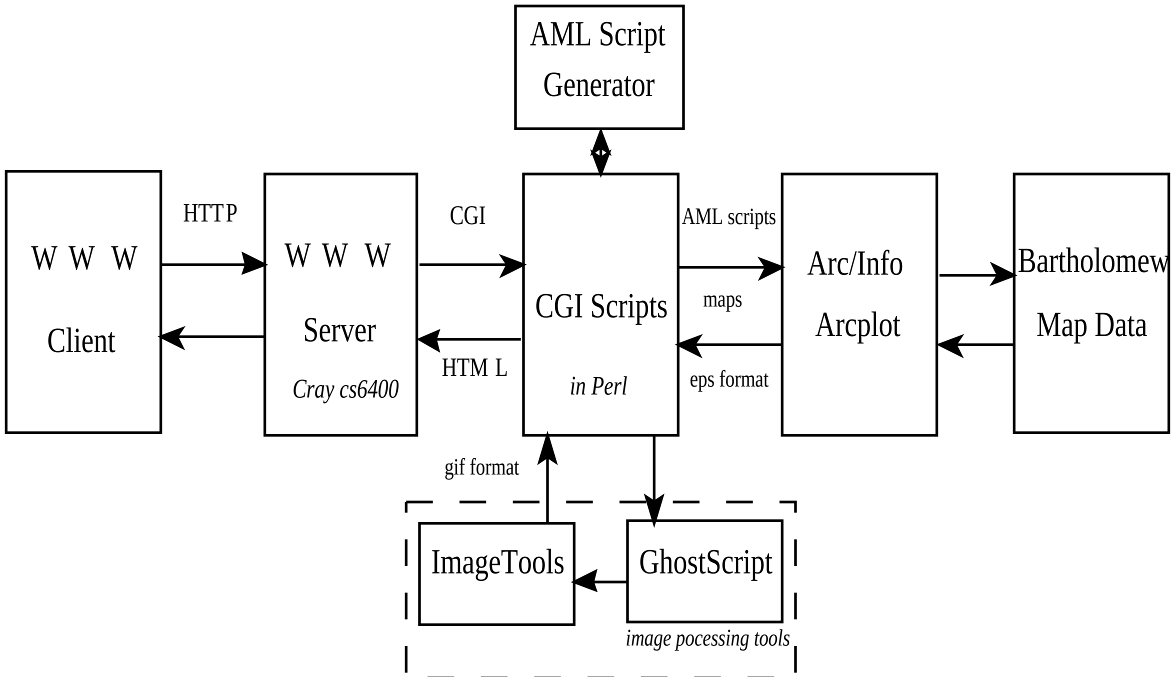

Figure 7: Outline of the Map Maker.

mented by a knowledge base to guide the users through use of the data.

6 Acknowledgements

The Knowledge-based Interface to National Data Sets (KINDS) project is funded by the Joint Information Systems Committee (New Technology Initiative NT-107).

Bartholomew digital data is Copyright ©Bartholomew and available to UK academics from MIDAS under the terms of the CHEST license agreement. We wish to thank Tim Rideout of Bartholomew for his support.

ESRI, Arc/Info, Arcplot and ARC Macro Language are registered trademarks of the Environmental Systems Research Institute, Inc., Redlands CA, USA.

7 References

Andresen, D; Carver, L; Dolin, R; Goodchild, M et al (1995) The WWW prototype of the Alexandria digital library.
http://alexandria.sdc.ucsb.edu/public-documents/papers/jap-paper/

Berners-Lee, T (1993) The WWW initiative and HTTP, HTML, etc. CERN WWW Documentation. http://www.w3.org/

Boston, T. and Stockwell, D. (1994) Interactive species distribution, mapping and modelling using World Wide Web. Proceedings of the Second International WWW Conference 94, Chicago, USA.
http://kaos.erin.gov.au/database/WWW-Fall94/species_paper.html

CIESIN (1995) *CIESIN Initiative*. http://www.ciesin.org/about/init.html

Davis, C and Medyckyj-Scott, D (1994) GIS usability: recommendations based on the user's view. *International Journal of Geographical Information Systems*, vol.8, no.2, pp. 175-189.

Frew, J; Carver, L; Fischer, C; Goodchild, M; Larsgaard, M; Smith, T and Zheng, Q (1995)
The Alexandria rapid prototype: building a digital library for spatial information. http://www.esri.com/resources/userconf/proc95/to300/p255.html

Fotheringham, A S and Rogerson, P A (1993) GIS and spatial analytical problems. International Journal of Geographical Information Systems, vol.7, no.1, pp. 3-19.

GENIE Project (1994). http://www-genie.lut.ac.uk/info.html

Li, C S; Kitmitto, K and Cole, K (1995) Developing a WWW interface to Arc/Info. Proceedings of ESRI'95 (UK) user conference, Nottingham University, Sept 14-15, 1995.

Li, C S; Moss, A et al (1995) Access Large and Complex Datasets via WWW. Proceedings of NTTS '95, Nov 20-22, 1995, Bonn, Germany.

Massem, P (1994) Demo of the interface between Arc/Info and the web.
http://www.geo.ed.ac.uk/home/research/massem.html

McLaughlin, J. and Nichols, S. (1994) Developing a national spatial data infrastructure. *Journal of Surveying Engineering ASCE* 120 2 pp. 62-76.

McCool, R (1994) The Common Gateway Interface, *NCSA WWW Documentation*. http://hoohoo.ncsa.uiuc.edu/cgi/overview.html

McGranaghan, M (1991) Matching Representations of Geographic Locations. In Mark, D M and Franks A U (eds), *Cognitive and Linguistic Aspects of Geographic Space*, pp. 387-402.

Medyckyj-Scott, D and Blades, M (1991) Cognitive representations of space in the design and use of geographical information systems. *People and Computer VI*, British Computer Society Conference Series, 1991, pp. 421-434.

Petch, J; et al (1995) The KINDS project. Proceedings of NTTS '95, Nov 20-22, 1995, Bonn, Germany.

Petch, J; Moss, A; Johnston, A and Yip, J (1995) Spatial data services: analysis of user needs. (to appear)

Plewe, B (1994) The GeoWeb project. Proceedings of the Second International WWW Conference 94, Chicago, USA.
http://wings.buffalo.edu/~plewe/paperwww.html

Raper, J. and Green, N. (1992) Teaching the principles of GIS: lessons from the GISTutor project International Journal of Geographical Information Systems vol.6, no.4 pp. 279-290.

Walker, D.R.F., Newman, I.A., Medyckyj-Scott, D.J. and Ruggles, C.L.N. (1992) A system for identifying data sets for GIS users. *International Journal of Geographical Information Systems* vol.6, no.6 pp. 511-527 .