Veterinaria Italiana, **43** (3), 709-721

Veterinary Web-geographic information systems: what's the point and what's involved?

William Wint

Summary

Geographic information systems (GIS) in general, and Web-based GIS in particular, are changing very rapidly because of the simultaneous increases in and sophistication of software, processing power, data storage capacity and available bandwidth. The spread of globalisation, with its consequent impact on trade, information exchange networks and emerging diseases, has meant that the demand for Web-based GIS is exploding. This has been compounded by the comparative speed with which basic sites can now be constructed. An active data site with a GIS element is a must-have for all self-respecting data-rich projects. There is a wide range of issues which should be considered before a GIS website can be launched – its function and appearance, its content and audience, its maintenance and stability and the implementation. There are also issues of technical complexity and data formats, levels of access, confidentiality and accreditation, or quality control and data validation, all of which must be addressed if a site is to be both reliable and effective. These and other topics are considered in some detail, with examples from around the Net.

Keywords

Animal health, Archives, Geographic information system, Spatial data, Website.

I sistemi informativi geografici veterinari sul Web: qual è il punto e che cosa comporta?

Riassunto

Il sistema informativo geografico in generale, e più specificatamente quello in rete (Web), si sta evolvendo molto velocemente grazie al simultaneo sviluppo di nuovi software molto sofisticati, di potenti processi di elaborazione dati, all'aumento della capacità di stoccaggio dei dati, nonché alla diffusione della banda larga in internet. La crescente globalizzazione, con il conseguente impatto sugli scambi, reti di scambi di informazione di dati e patologie emergenti, ha creato l'urgenza di un sistema GIS nel Web. Questo grazie anche alla facilità con cui un sito oggi può essere creato. Per un sito accreditato di raccolta e studio dati che si rispetti è oramai doverosa l'integrazione con un sistema GIS. C'è una gran quantità di problemi organizzativi di cui bisogna tener conto prima di immettere un sito GIS in rete-la sua funzionalità e l'aspetto grafico, i suoi contenuti ed eventuali fruitori, la sua manutenzione e stabilità nonché l'implementazione dei dati. Ci sono inoltre problemi di natura tecnica come il formato dei dati, i vari livelli di accesso, riservatezza e credibilità, nonché controlli di qualità e validazione dei dati; si deve mirare a tutto questo per far si che un sito sia affidabile ed efficiente. Questi ed altri argomenti sono considerati in dettaglio con esempi specifici sulla rete.

Parole chiave

Archivi, Dati spaziali, Sanità animale, Sistema informativo geografico, Siti Web.

Environmental Research Group Oxford, Department of Zoology, South Parks Road, Oxford OX1 3PS, United Kingdom
william.wint@zoo.ox.ac.uk

Introduction

Web-based geographic information systems (GIS) depend on the accessibility of several critical resources, namely: software, data, Web, hardware to accommodate them and, more importantly, the skills to design, construct, populate and operate them. For those of us who are immersed in the field, it is all too easy to assume that these elements are and have always been available. This is not so.

Twenty-six years ago, in 1981, GIS were the preserve of institutional computer centres or departmental minicomputers; Apple was five years old; Microsoft had 128 employees; and a PC's storage capacity was measured in Kilobytes (about equivalent to a three-page Word 2003 document). Twenty years ago PC ArcInfo™ was first released, Idrisi was in beta development; PC storage capacity was measured in Megabytes and file transfer between computers was largely limited to file transfer protocol (ftp) between academic or government mainframes linked by the joint academic network (JANET). Email was a novelty and the World Wide Web was four years from birth (20). By the mid-nineties, email was used widely (at least in academia), PCs could hold 100 MB of data, Windows for Workgroups had been around for two or three years, and data exchange was a regularly used facility within institutions with the required physical wiring. A mere ten years later, GIS programmes are mainstream applications, with entire corporate and academic departments devoted to their use and development; data storage capacity is expressed in Terabytes, PCs can be less than US$500, data transfer rates are measured in MB per second and broadband is now common place, even in the low-income countries.

As a result, many of the factors that traditionally limited large-scale geographic analysis are dissolving: the software available has diversified to include the monolithic and the open source, the specialist and the generalist; it has matured to cope with and convert between the variety of data formats (raster, point, vector, database); the suites include the range of utilities required for reliable analyses (mapping, display, data manipulation, statistics); a current mid-range PC designed for graphics intensive games will readily process all but the most complex spatial statistics, and internet connections and servers that sustain movie and iPod downloads, or search engines serving 70 billion hits a month can easily route a few satellite images to a researcher's laptop. Furthermore, the plethora of Web-authoring tools replete with extensions and add-ons and the variety of remote hosting solutions have brought the construction of a fairly sophisticated website within reach of the amateur with a few hours to spare.

In another ten years, it is hard to imagine that there will be any technical information technology constraints to the construction of GIS-based websites. Indeed, GIS and mapping is already ubiquitous – if data are only presented in tables they are incomplete – a truly remarkable impact for a tool that has been widely available for less than a decade.

There remain, of course, a number of limitations. Reliable geographic analyses depend on reliable geographic data, and because the whole ethos of GIS is founded on the combination of layers of information of many types and from many sources, it is difficult to ensure the availability of compatible data from a wide range of suppliers. This is particularly true of epidemiological information from the low-income countries, where government resources are scarce and data gathering or surveillance is a low priority, and is often left to non-governmental organisations (NGOs) or externally funded projects that are often short-term. The other constraints that no longer apply in richer nations (hardware, software, network capacity) are also still very much in evidence in the poorer ones. This situation may improve, however, as wireless and satellite communication become more widespread, and as the move towards using open source utilities accelerates.

Pervasive barriers to the effective development of many GIS-based websites are those raised in the name of confidentiality or data protection.

This is especially true of those dealing with the health of animals, plants or humans, and therefore with potentially dramatic impacts on trade and local economies. Unlike the constraints so far mentioned, this is primarily bureaucratic, and thus likely to proliferate rather than fade away. Data ownership and intellectual property are also often cited as reasons to restrict data dissemination, especially by academics who frequently withhold data from the public domain until the information is formally published (see below). Fortunately there are increasing calls for disease data to be made available more freely (2).

This rapidly developing context defines the supply of information that populates veterinary GIS and its dependent websites. The demand for the information is not governed by the same limitations. Certainly the planners are increasingly aware of the potential uses of geographic and mapped data: a picture is worth a thousand words and decision-makers increasingly require that information is mapped. The veterinary and research communities that utilise and adapt GIS analyses are expanding and the demand from government and media for disease-related information is raising the requirement for easily accessible centralised disease information repositories, pre-digested summaries, and, increasingly, early warning systems. All depend on networked data exchange and standardised data processing: in short, various forms of (veterinary) GIS websites.

The following overview looks at the issues involved in constructing a veterinary GIS website, from the point of view of both supply and demand – webmaster and user. It attempts to identify the cardinal sins to be avoided and to describe the decisions (and compromises) that inevitably need to be made before a website is launched. These will include its function and appearance, its content and its audience, its implementation and its maintenance. There are also issues of, for example, technical complexity and data formats, levels of access, confidentiality and accreditation, or quality control and data validation, all of which must be addressed if a site is to be both reliable and effective. These and other topics are considered in some detail, with illustrative examples.

Reasons for a veterinary geographic information system website

Veterinary GIS in general, and websites in particular, have often been criticised as technology in search of an application, and there is little doubt that some elements of this perception have been justified because little reliable data have been available to populate the sites. In recent years, however, more veterinary disease data has been made available on official websites and the public demand for information on the high-profile emerging diseases (avian flu, bovine tuberculosis, bluetongue, foot and mouth disease) has pump-primed an explosion of data-rich early warning and disease risk assessment projects. Such projects have usually produced websites to showcase their wares but which more often describe project objectives and anticipated outputs. Frequently the actual results are less evident, or are described as 'for internal use only'. Such websites might be more appropriately named intranet sites, as their content often fails to reach the public domain. They do, however, serve to advertise data and hopefully, at the very least, give the contact details of someone to approach to acquire the geo-referenced data. Sadly, even this is often not the case, as the specified contact has moved on and the email account disabled, one of the cardinal sins set out in Table I.

There are, however, a variety of potential justifications for producing websites that are of wider use, rather than just being (the first) evidence that a project exists. Some are aimed at interested users with little GIS technical experience, which tend to produce graphic and textual output summarising GIS work (for example, avian flu maps from the BBC and the United Kingdom Ordnance Survey) (1, 16), whilst others tend towards geo-referenced disease data, attribute provision (7, 10, 21), or multidisciplinary GIS format data (8) for use by GIS professionals.

Table I
Cardinal sins compiled from a straw poll of geographic information systems colleagues

Cardinal sins from a user's perspective
▪ Works only on Internet Explorer (requires a plug-in
▪ Site with objectives but no outputs
▪ Data stated to be available for download but not actually accessible, or link to metadata only
▪ Broken internal links
▪ Broken map servers
▪ Data unusable– i.e. wrong bounds/codes/resolution in metadata
▪ Use of rare or esoteric data formats and projections
▪ Missing metadata (contact address unavailable)
▪ Slow response to registration requests
▪ Impenetrable page or menu hierarchy
▪ Pages available through search pages but not menus
▪ Slow display or unreliable basic functionality (especially of map servers)
▪ 'Coming soon' never materialises
▪ Infrequent updating
▪ 'Under construction' never disappears (old versions of revised sites remain accessible/visible to search engines

From the producers' point of view, publicity, dissemination and presentation of results are frequently a major component of the outputs required by funding bodies and they do serve to alert users to potentially valuable information. Once actual results become available as, for example, risk maps, disease distribution data or epidemiologically relevant data archives for research use, the graphic descriptions and overviews can be used as a basis for reports and proposals, without recourse to the actual data. An extension of this idea is the provision of interactive maps that depend on map-server technology such as ArcIMS™, which allows users to define maps tailor-made for their own purposes. Usually, however, these sites can only provide graphic or printed output and cannot be used as a source of actual data. Exceptions to this rule are the satellite imagery download sites (14, 18) and multidisciplinary data clearing houses (6, 11) essential for many animal disease modelling methods which are developing increasingly interactive ways to help the user select data to download.

A valuable and increasing use of GIS in veterinary epidemiology is the production of risk maps which can be disseminated readily over the Web. At present, these are surprisingly few and far between, partly because GIS technicians tend not to be modellers and modellers are often less spatially aware than might generally be assumed. The promotion of GIS-based risk models on the internet will hopefully encourage a cross fertilisation of the two disciplines. There is also an awareness of the possible dangers of making risk maps widely available, particularly the potential liabilities.

Ancillary GIS website content is often as useful as the primary information – supporting tutorials, helpfiles, descriptions of methods and relevant tools or utilities – often helps to bring what can be quite technical information to an interested lay audience. Links to other relevant sites are another potentially valuable use of Web space – particularly as a substantial links page can dramatically raise the visibility of a site to the major search engines.

Purists would probably define the most valuable function for a full-blooded GIS website as the provision of a multidisciplinary GIS format data archive for use by other researchers and projects who are fully GIS literate. Such sites can save others significant time and resources by ensuring that different data sources and types are compatible and comparable and, as far as possible, error free. Archives are useful not only to disseminate data from mature projects, but are an efficient means of providing an infrastructure for working data collection and exchange between project members (6). These can ensure that everyone within a specific project is 'singing from the same hymn sheet' using, for example, a single set of human population figures as the denominator for prevalence calculations. Such working website archives are also a way of reducing data duplication amongst disparate project members.

Finally, data reporting and collection is increasingly achieved by user populated websites relying on the submission of observations or regular records to centralised databases, managed and interpreted by GIS professionals. Examples of these are presented elsewhere in this volume (4, 13) and so will not be considered here.

It should perhaps be stressed that, like the enabling technology, the drivers for website production are changing rapidly. More and more funding agencies and institutions are requiring that the actual data from their projects are put in the public domain, though the lead in this transformation of priorities is being taken by the medical (19) rather than veterinary world. Initiatives like GISAID, recently proposed for the pre-publication sharing of avian flu data are also promoting such changes in attitude towards data sharing (2). This trend must be encouraged: veterinary disease data in the public domain are a rare commodity in comparison to the huge expansion in global datasets covering environmental, demographic and climatic information that are now becoming available. There is a danger that this discourages research efforts in the veterinary area in favour of topics with more accessible information, thereby hindering the development of much needed spatial analysis methodology

User identification

Which of these various potential roles are assigned to a GIS website depends not only on the aspirations and requirements of the producer, but also very significantly on the needs, skills and technical resources available to the users. Levels of GIS, map and computer literacy are very variable within any professional community and will, to a large extent, govern the ability of users to access and interpret veterinary GIS information. Because GIS analysis requires specialist and often multidisciplinary data processing skills (data formats, projections, statistics, etc.), data archives and clearing houses tend to be the territory of GIS professionals. These are still a fairly rare commodity and most users are not able to use geographic data in their raw format, and so rely on graphic, textual or interactively produced maps. There is no point therefore in providing GIS format data files to non-specialist users, though equally, there is little mileage in only providing graphics format output like bitmaps and document figures to GIS technicians.

Perhaps most importantly, the location and diversity of its user community will be critical determinants of the design and 'feel' of a GIS website, as well as influencing whether a site should be multilingual, and determining the browsers and operating systems with which it needs to be compatible. Web developers tend to use the latest software and the fastest machines and may forget that even the major browsers can produce very different displays from the same set of instructions. Backward compatibility may be a significant problem with a global user base using software and operating system versions that differ by five years or more, and the wider the potential audience, the more attention that should be paid to testing on different client configurations.

The users' technical skills will not only determine the type of primary content, but also the level and type of supporting information needed. Lay users may well require interpretive materials, a measure of technical support and pre-packaged outputs, but will not need the specialist methodological information and technical metadata that professionals will require.

The users' own skills and hardware capabilities may also tend to dictate the server/client balance – i.e. how much of the website processing is done on the machines hosting the website and how much is left to the user. Map servers may be more appropriate to supply mapped materials to the less GIS-literate. Sites with many images, or involving a lot of data extraction, downloading or map production will, for example, place large demands on the server, which is likely to be tested more if there are many users with broadband access. Graphics and large downloads will deter (or frustrate) users with high contention ratio broadband, or connections that are frequently interrupted. It should be emphasised that these are not limited to low-income countries and institutional websites often assume a far higher network speed (perhaps more appropriate to intra- rather than inter-net) than may be available beyond their own firewalls.

The issue of bandwidth against file size is one of a number other potential 'conflicts of interest' between the needs of the users and those of the producers; file access and data security; liability and confidentiality; site complexity and ease of use, can all require substantial compromises if both supply and demand is to be satisfied. These issues will be addressed below.

Complexity and content

Once the users have been characterised and the major objectives of the website defined, there are many decisions to be made which affect the resources required to set up and maintain a website, as well as defining the type and security of the information it contains.

Perhaps the first of these is to decide the proposed site's anticipated shelf-life and whether it is destined to outlast the funding cycle which made it possible. This is especially important for data archives and for project outputs – which can often be at their most useful for the period immediately after a project has been completed, and the resources needed to maintain a site withdrawn. This is when the information is still new enough to be interesting to a wide audience and, if possible, arrangements should be made for continued hosting past project wind-up dates.

There are some sites that have longer currency than others: for example, epidemiological data, methodology reviews or details of techniques are likely to be valuable for much longer than links to related (and equally defunct) sites or to those that were set up at the beginning of a project detailing the many ambitious plans and milestones, which though eventually realised, were never posted to the project site because the webmasters have moved to new jobs.

The range of possible website types is endless – from working blogs (Web logs or diaries), via sophisticated graphically intensive project descriptions and dissemination sites, to complex data archives with user-defined geographic selection, rapidly changing content and high-level security and user validation procedures. Data archives can contain working data to promote collaboration and exchange, 'beta version' output for comment and revision, or final research ready data for public dissemination, or of course a mixture of all these categories (6) (Fig. 1).

The technical resources used to construct sites are equally variable. At one end of the spectrum are the simplest tables with hyperlinks leading directly to downloadable files (9, 17), often written in free Web authoring packages and using server software provided within standard operating systems, such as Microsoft® Sharepoint. More complex are the hypertext pre-processor- (php) and sequential query language- (SQL) based sites running under Microsoft® Server, or the open source Apache. More complex yet are the map-served applications generated by expensive internet mapping software suites, such as Environmental Systems Research Institute (ESRI) ArcIMS™, or sophisticated utilities written in Web programming languages like Java, or relying on tools like ActiveX plugins or Flash Media content.

Perhaps most recently, the advent of Google™ Earth and the ability to reproduce ESRI ArcGIS™ maps from comparatively small data files (e.g. keyhole markup language or kml) that can be posted or emailed has opened up additional possibilities where the generic map content (backdrops, boundaries, populated places) is provided by central servers, and only the user overlays are needed to provide quite sophisticated and customisable displays, without the need for any geographic software installed. A widely used example is (again) the Google™ avian flu maps layer (12). This is very simple to set up (a single click in Windows Explorer), the data are supplied

Figure 1
The Emerging diseases in a changing European environment (EDEN) website illustrating a breadth of content

from a single source, so updating and maintenance is straightforward and given the fact that they are also entertaining to use, they are likely to spread rapidly and become a significant source of geographic data of all types.

The ongoing stabilisation of data formats promoted by the Open Spatial Consortium may also provide the opportunity for GIS websites to draw data and layers from a range of different archives rather than a single remote server, to be combined using free GIS visualisation tools in a single interactively defined output (3).

GIS websites do not, however, have to be all singing, all dancing multimedia extravaganzas in order to be useful. For data exchange or acquisition by professional GIS technicians, within small teams or projects, remote access or file transfer protocol (ftp), backed up by email communications and advisories can be sufficient for most purposes. These do tend, however, to get quite complex to administer if archive folder structures are very complex, if data content is very diverse or requires substantial descriptive content, or if there are many users. Data security is also a major issue. In addition, of course, the front-end designs of such sites are often rather utilitarian and user-friendliness rather limited. On the other hand, these are very easy to create and require very basic (often free) software. Equally simple are the increasingly widely available voice over internet protocol (VoIP, such as Skype™ or Google™ Talk) file transfers or third party file transfer facilities (e.g. yousendit.com) which can cope with files of up to 1 GB or more.

The relentless spread of the Web culture has also provided many options for website hosting. Whilst institutions and large projects usually have the capability to host websites on their own machines, the use of hosted websites on internet service provider (ISP) servers (e.g. AOL or Wanadoo), using referred domain names is becoming increasingly common for those sites that do not contain large data volumes (i.e. less than 100 MB) and consist of comparatively few pages. These have an advantage for those with less reliable internet access, as the major ISPs are usually stable and reliably accessible, and a connection to the Net is only needed to update the hosted site.

The choices for Web structure and content thus range from the very simple to the very complex. Whilst simple may be preferable in terms of resources and support required, there are at least two good reasons to select the complex option, as follows:

- to provide flexible (interactive user selected) graphical outputs
- to control access, thereby ensuring data security and confidentiality

This last topic is considered in the next section.

Access, security, confidentiality and liability

One of the main objectives for a veterinary GIS website may be project publicity and dissemination, generally measured in the number of page hits. Animal health information, especially regarding disease distribution, may however have severe (and negative) economic impacts through its potential effects on veterinary status and thereby exports, on levels of government control and on demand for products. Potential liability for the accuracy and reliability of any data or advice provided may not only necessitate attaching disclaimers to information provided, but may also mean that the number of users needs to be restricted or that access to certain sections of a website, or to specific datasets it contains, are limited to validated users. This is a particularly important issue if distributing data falls foul of national or international confidentiality and data protection laws. Even the maintenance of user registers may be affected by such legislation.

Information, especially that generated by research and development projects, may also be subject to jealously guarded intellectual property rights and the owners, if they can be persuaded to supply their data at all, may well impose draconian restrictions on access to their information, even if watertight data sharing agreements are in place. Despite the availability of effective firewalls and website security, there remains a widespread perception that once data is posted on a website, it is available to anyone and everyone who wants it. As a result, perhaps the greatest hurdle to overcome, as far as data provision is concerned, is to negotiate agreed access permissions and distribution rules with the data suppliers.

These concerns are especially relevant to processed GIS data. Many projects acquire data, even for formally approved use, through informal contacts between technicians, or at the other end of the spectrum by signing fearsome and legally binding confidentiality agreements. Such data may be used only as one of several inputs to a final processed output and it is often very unclear whether redistribution of such products is governed by the same restrictions as were the original confidential raw data sources. Rigorously applied attribution and acknowledgement of original sources is essential; the use of licensing protocols, such as the GNU General Public License or the Creative Common Access License (15) and detailed monitoring of user access and downloads can help address these problems.

Such issues raise a number of practical questions – particularly how to manipulate levels of access to sites whilst controlling access to data within them. Maximising access and user visibility tends to be more of a challenge than is the reverse. Though traffic levels, especially to graphic or server-intensive sites, may be limited by available server capacity and user bandwidth, the majority of sites need to take positive action to increase their hit rate, which is increasingly dictated by their visibility to the major search engines. Appropriate keywords, links to and on sites hosted at authority domains, incorporation of links pages, regular updating, measured site expansion, and an established page history all help to enhance page rankings.

It seems likely that the number of site hits, page views, registered users or the page rankings calculated by the search engines will soon become as important to project and career assessment as citation indices and publication numbers are currently. Site log analysers like Analog or traffic monitors like Google™ Analytics already provide putatively independent site statistics and there are browser toolbars which incorporate rankings of viewed pages. Presumably there will soon be a system which provides certified traffic statistics for external site evaluation.

Restricting access is more of a technical process. Simple password protection of pages or files will deter many casual attempts to access confidential data, though is readily circumvented by those who can spare the time, or the few tens of dollars necessary to buy cheap password cracker software packages. User registration and validation with multi-level user and data-specific access rights (16, 21) requires substantially more complex website architecture and data archive structures, where data are catalogued using, for example, SQL databases wherein all data files are assigned access group codes relating to user rights. The data files are usually stored outside the website folder structure to prevent unauthorised folder listing and browsing and hyperlinks point to the SQL database, not the data files themselves.

There are substantial overheads to data security – not only the level of programming needed to build, maintain and add to a website or archive, but also the time needed to administer user validation procedures and requests (6) (Fig. 2), and, particularly for very busy sites or low bandwidth users, there may be significant performance penalties associated with the logging and monitoring functions that are part of the security functionality.

Data archives: content, format, metadata and updating

Epidemiology and GIS both rely heavily on multidisciplinary analyses and so are characterised by the use and combination of

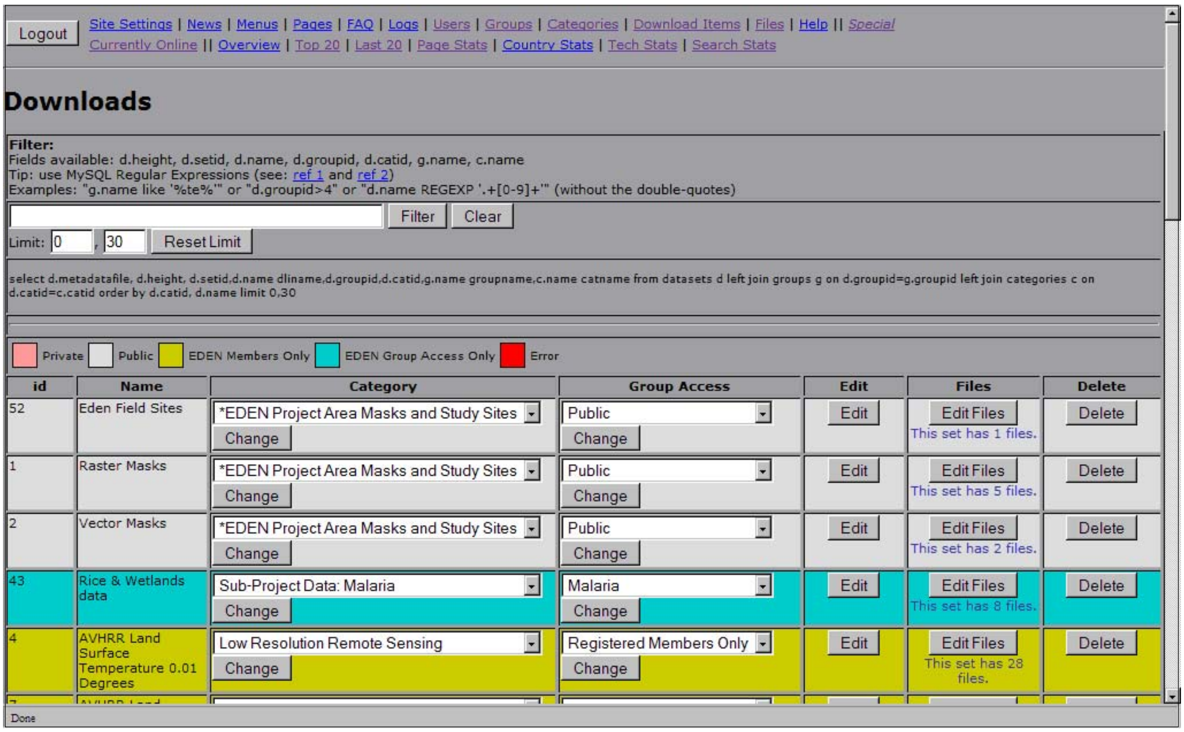

Figure 2
Emerging diseases in a changing European environment (EDEN) data site – flexible administration of data access

information from a very wide range of sources. One of the major functions of a GIS website is to process available data to provide not only standardised outputs, but also standardised inputs for collaborators to use in their analyses. Failure to do this can lead to major discrepancies in research results – for example, using different sets of human population estimates, perhaps derived from different resolutions, can lead to wildly disparate calculated disease prevalence figures. Polygons from different data sources may not have the same estimated areas because their projection or node density varies or because different versions cover slightly different areas.

The obvious geographic criteria to standardise are bounds, projections (including datum and spheroid), resolution or scale. Perhaps less obvious are polygon boundaries (especially of the ever changing administrative units), areas, names, analysis masks and, if possible, reference years for supporting data layers like populations, land use or environmental time series and averages.

Whilst standardisation is a must, limitation to a single set of approved criteria may impose too many restrictions on the users, and it may therefore be most effective to have more than one set of standards – particularly for such things as projections or data file formats. Thus, standard projections for example, should include geographics to ensure compatibility with the majority of public domain geo-referenced data, but also one of the metric projections, to facilitate manipulations based on area. There is also a case for adopting universal transverse mercator (UTM) as a standard projection, because it is so widely used by GPS units, though with the caveat that it is all too easy to misdefine the UTM zone, thereby invalidating all the geo-referencing.

Choice of file formats depends, as ever, on user capabilities. For the specialist, most formats will be usable, whilst for a wider appeal, file types are likely to be dictated by the dominant software of the moment. These may differ from country to country, – but are most often Adobe®, Microsoft® and ESRI file types or ones that can be imported or exported by their software. Thus documents tend to be *pdf* (Adobe portable document format), *doc* (Microsoft® Office document) or less frequently *rtf* (rich text format); graphic images tend to be *jpg* (joint photographic experts group), *tif* (tagged image file format) or *gif* (CompuServe® graphics interchange format); numerical information such as attribute data is usually supplied as *dbf* (Dbase format), delimited text files, or formats such as Excel or Access that can readily import or export them. Raster imagery tends to be raw binary, hierarchical data format (HDF), ERDAS® Imagine or ESRI grids, though the number of files in the latter can cause problems, and less complicated formats such as Arc™ Export (*e00*) or geoTiff are also widely used. Very large or long established data archives dealing with specific topics may provide files in their own format rather than the generic ones just mentioned (5). Vector files are almost universally distributed ESRI *e00* or *shp* (shape files).

A further factor to bear in mind is the level of compression the selected data formats can support for distribution in either their native format or within zip or rar archives. Heavily compressed files are very good for downloads but can cause real problems when they are subsequently extracted and the size of their native format reasserts itself. In some instances (e.g. ESRI binaries, or ASCII text files), compression ratios of 100:1 can mean that download of 10 MB suddenly occupy a gigabyte of valuable disk space.

A key component to all Web-posted data is the provision of supporting metadata (for data files) and reference sources or links for graphics and textual descriptions. Preparing and maintaining these can be a major overhead to data suppliers and archive managers alike, but are essential to provide not only the technical details that allow geographic and numerical data to be used, but also current details of a contact to whom questions about data or their availability can be addressed. At the very least, readme files should accompany data sets, giving salient details and sources. However, particularly for geographic data, metadata compilation and definition is an industry in itself, and the standard contents that are required to conform to international data standards are many and various (though the Open Standards Consortium may simplify this). Users' lives can be simplified by using metadata compilation tools, or by reducing the number of required items, though it is often the case that the more minimalist is the metadata description the more user questions are generated for the supplier to answer. The combination of a reduced metadataset per data file and a more detailed data summary or overview relating to multiple data files is a solution adopted by many. It is, however, essential to be certain that the ancillary data are accurate – as incorrect parameters (a cardinal sin, Table I) can render downloaded data unusable.

User needs and cardinal sins

The foregoing discussions have focussed on the supply-side choices that need to be made by a website designer or manager. Whilst it has been stressed that these choices must be influenced by the assumed user requirements, there are certain factors that are of particular concern to the user. Most of these relate to the list of cardinal sins set out in Table I.

Users visit websites to locate information, download data or, less frequently, to provide information to a centralised data archive, so it is important to make locating their target pages or files as painless as possible. Search pages and site maps can help achieve this objective, though nested hierarchical links or menus are more often used. The three click rule is paramount – it should take no more than three clicks to definitely locate required information (though it might take one or two more to actually acquire it). Though the rise of Google™ and its peers have reduced the need for complex search pages, it has made the navigation structure all the more important as a visitor's first view of a site is less and less likely to be its home page, and more likely to be at the location of the searched for item.

It is also important to make sure that Web pages don't contain too much information – and that it isn't necessary to move down or across a page to locate something – screen resolution is therefore a significant issue, as it can vary radically from one user to another, and imposes strict limits on graphic and font sizes. Users usually have lower screen resolutions than Web designers.

Key to user approval is site stability and display speed. This may require Web designers to trade complexity for speed, especially on heavily visited sites or those aimed at low bandwidth users. Slow downloads may be blamed on connection speed, but long delays in display or page loading, though also possibly due to poor bandwidth, is usually ascribed to poor design. Preview graphics, thumbnail displays, text only pages, optional implementation of multimedia and animated display, and using new pages or tabs to open slow loading parts of a website can help to reduce user aggravation.

It is essential to ensure that Web content and links which lead to information is as accurate and current as possible, which requires regular maintenance and updating – and is an overhead that many sites do not appear to pay. Pointers to data or resources that are 'in progress' or lead only to descriptions or contact names, are a frequent occurrence on GIS data sites and are at best misleading, if not actively deceptive. This is especially the case if answers from contacts supplied are slow to materialise or are referred onwards for responses – again a sign of insufficient support and maintenance. Sites which provide graphic output of distributions (via images or map servers) only to state the underlying data are 'for internal use only' also run the risk of alienating potential users.

Conclusion

GIS websites are in their infancy, but are growing up fast, as the technological and infrastructural constraints are overcome and the volume of available data proliferates. As a result, sites with maps are now commonplace. Not all of them, however, are produced with the priorities of user driving the design and content, a situation which should not be encouraged. Perhaps most importantly, GIS websites are not something to 'fire and forget', but rather require constant attention and development if they are to fulfil their huge potential.

Acknowledgements

My thanks are due to Andy Tatem and Simon Hay of the Malaria Atlas Project (www.map.zoo.ox.ac.uk); to Neil Alexander at the Environmental Research Group Oxford (ERGO) for providing valuable input and constructive criticism; and to my wife Sarah Wint for her patience in correcting my grammar and proof-reading manuscripts.

Grant support

The author received financial support for the Food and Agriculture Organization of the United Nations, the United Kingdom Department for the Environment, Food and Rural Affairs, and from European Union Grant24 GOCE-2003-010284 EDEN (Emerging diseases in a changing European environment). The contents of this publication are the sole responsibility of the author and do not reflect the views of any of the funding bodies.

References

1. British Broadcasting Corporation (BBC) 2006. How bird flu has spread (news.bbc.co.uk/1/shared/spl/hi/world/05/bird_flu_map/html/1.stm accessed on 9 July 2007).
2. Butler D. 2006. Disease surveillance needs a revolution. *Nature*, **440** (7080), 6-7.
3. Butler D. 2006. Virtual globes: the web-wide world. *Nature*, **439** (7078), 776-778.
4. Calistri P., Savini L., Ippoliti C. & Conte A. 2006. Web-based geographic information system tools and international surveillance network for bluetongue in the Balkans and south-east Europe (East-BTnet). *In* Geographic information systems (P. Calistri, A. Conte, J.E. Freier & M.P. Ward, eds). *Vet Ital*, **43** (3), 739-744.
5. Climate Research Unit (CRU) 2006. CRU TS 2.1 datasets (www.cru.uea.ac.uk/~timm/grid/CRU_TS_2_1.html accessed on 9 July 2007.
6. European Commission 2006. Emerging diseases in a changing European environment (EDEN) data website (ergodd.zoo.ox.ac.uk/eden accessed on 9 July 2007).
7. European Commission 2006. European Union Statistics (EUROSTAT) (ec.europa.eu/eurostat accessed on 9 July 2007).
8. European Environment Agency (EEA) 2006. Overview of EEA products (www.eea.europa.eu/products accessed on 9 July 2007).
9. Food and Agriculture Organization (FAO) 2006. FAO Pro-poor livestock policy initiative (PPLPI) 2006. Inter-Governmental Authority on Development (IGAD) livestock initiative working website (ergodd.zoo.ox.ac.uk/igadweb accessed on 9 July 2007).
10. Food and Agriculture Organization (FAO) 2006. FAOSTAT, FAO statistics archive (faostat.fao.org/default.aspx accessed on 9 July 2007).
11. Food and Agriculture Organization (FAO) 2006. GeoNetwork 2006 (www.fao.org/geonetwork/srv/en/main.home accessed on 9 July 2007).
12. Google Earth 2006. Avian flu maps (declanbutler.info/Flumaps1/avianflu.html accessed on 9 July 2007).
13. Maroney S.A., McCool M.J., Geter K.D. & James A.M. 2006. The evolution of internet-based map server applications in United States Department of Agriculture, Veterinary Services. *In* Geographic information systems (P. Calistri, A. Conte, J.E. Freier & M.P. Ward, eds). *Vet Ital*, **43** (3), 723.730.
14. National Aeronautics and Space Administration (NASA) Distributed Active Archive Center (DAAC) 2006. Goddard Earth Sciences, Data and Information Services Center (daac.gsfc.nasa.gov accessed on 9 July 2007).
15. Public Library of Science (PLOS) 2006. Open Access License, Summary of the Creative Commons Attribution License (www.plos.org/journals/license.html accessed on 9 July 2007).
16. United Kingdom Ordnance Survey 2006. MasterMap (www.ordnancesurvey.co.uk/oswebsite/products/osmastermap/index.html accessed on 11 July 2007).

17. United Nations Environment Programme (UNEP) 2006. Division of early warning and assessment (DEWA) Global resource information database (GRID) data archives (www.grid.unep.ch/data/index.php accessed on 9 July 2007).
18. University of Maryland 2006. Global land cover facility (www.landcover.org/index.shtml accessed on 9 July 2007).
19. Wellcome Trust 2006. Malaria atlas project (www.map.ox.ac.uk accessed on 9 July 2007).
20. Wikipedia 2006. History of the internet (en.wikipedia.org/wiki/History_of_the_Internet accessed on 9 July 2007).
21. World Organisation for Animal Health (OIE: Office International des Épizooties) 2006. Handistatus II, Prototype (www.oie.int/hs2/report.asp?lang=en accessed on 9 July 2007).