INSPEL 36(2002)1, pp. 22-48

GEOSPATIAL DATA ACCESS: CAN WE MANAGE TO SHIFT?[*]

By Jan Smits

Our new earth

In his book *Snow Crash* Neal Stephenson portrays a virtual librarian. This is a digital construct, which looks and acts like a virtual human librarian, as this is easier to communicate with than an icon or any other concentrated digital display. The library he works with is a hyper library, i.e. all the data he has access to are interrelated, either by themselves or because of historical use. The librarian works with Hyper Cards and contains a self-educating algorithm, i.e. he has "the innate ability to learn from experience[1]". When asked who wrote him the librarian answers:

"I was … coded … by a researcher at the Library of Congress who taught himself how to code. … He devoted himself to the common problem of sifting through vast amounts of irrelevant detail in order to find significant gems of information". [2]

The most relevant source of information for us in this book is a piece of software called 'Earth', which "keeps track of every bit of spatial information that it owns – all the maps, weather data, architectural plans, and satellite surveillance stuff". When we take this one step further we may also add to these geospatial data all locational data, which is part of analogue and digital alphanumerical objects and databases, keeping in mind that a lot of spatial phenomena, including human activity, are constantly monitored[3].

---

[*] Paper presented at the 67th IFLA General Conference, Boston USA, August, 16-25 2001. Section on Geography and Map Libraries.

[1] This may result among other things in the filing of profiles which are constructed of previously posed queries and answers, i.e. FAQ's and FGA's, but the profiles may also include topological or contextual intelligence.

[2] Stephenson, Neal (1992). *Snow Crash*. – London : ROC, p. 101

[3] The amount of monitored data with a spatial attribute is overwhelming. To give but a few examples: satellites constantly spew data concerning the physical environment for e.g. weather-forecasts and the Global Change project and keep track of movements in the air and on the water; governments monitor traffic and other movements and keep track of planning and building; censuses are taken at intervals by most countries; biologists recount at intervals biological presence; companies keep large databases of customer-behavior; the world is almost spammed with web/video-cams; etc.

Thus one of the possible futures toward which we are trending or are unconsciously becoming part of is sketched.

What we have to assess is whether we are (going to be) able to work with access tools -whether existing, developing, or probabilistic- which make it possible to find the 'gems of information' our customers are looking for. In this assessment we have to keep in mind that we still mainly deal with Document-like Objects (DLO)[4], but that in the future we will more and more have to locate first the relevant data and then the software to create a legible file and/or image and possibly a tactile image. This transition from DLOs to databases[5] will take quite some time and depends partly on the availability of the latter in the public domain, i.e. in libraries, archives and the like accessible for private use. Should the data not be available in large quantities in the public domain than standards for its filing, retrieval and use will be market conforming and hopefully interoperable[6] and available to the public at large.

Towards a new paradigm

---

4 Though the term is much in use on Internet sites concerning the documentation of Internet sources it is hard to find a definition thereof. After reading through some 30 Internet sources I did not find an adequate definition. I would define a DLO as a unit of information that carries distinct descriptive data, which can be used in a formal description. The following quotes may lead to a better understanding of the nature of DLOs:

Document-like objects which share the characteristic that they can, but need not be, adequately represented in a print format, examples being pure text, textual images, text with printable illustrations, and photographs.

POCKLEY, Simon (1997). *Killing the duck to keep the quack.*
<http://www.cinemedia.net/FOD/FOD0055.html> [Accessed: 11 January 2001]

Non-document-like objects, on the other hand, include such resources as virtual experiences, databases (including ones that generate document-like outputs), business graphics, CAD/CAM or geographic information generated from database values, and interactive applications which might have different content for each user. In the context of image discovery, these sources do not "contain" images as much as they "generate" images. The images they generate may be described as fixed document-like objects, but the metadata required to describe them (the systems doing the generating) are distinct.

WEIBEL, Stuart, and Eric Miller (1996). *Image Description on the Internet*. A Summary of the CNI/OCLC Image Metadata Workshop, September 24 - 25, 1996, Dublin, Ohio. In: D-Lib Magazine, ISSN 1082-9873, January 1997.
<http://www.dlib.org/dlib/january97/oclc/01weibel.html> [Accessed: 11 January 2001]

5 Although R2V (raster-to-vector conversion) and satellite imaging of Land Use/Land cover classification are well established, the latter are mainly in the form of DLOs and not as databases. I would define a database as a collection or BLOB (Binary Large Object) of distinct data which are retrievable through an underlying structured principle, e.g. geographic co-ordinates, the periodic table of elements in chemistry, the genome sequence, etc.

6 Interoperability (as defined by the FGDC): A condition that exists when the distinctions between information systems are not a barrier to accomplishing a task that spans multiple systems.
Nerbert, Douglas D. (2000). *Z39.50 application profile for geospatial metadata or "GEO"*
<http://www.blueangeltech.com/Standards/GeoProfile/geo22.htm> [Accessed: 11 January 2001]

We are at the moment going through what might be called a paradigm shift. Kuhn used the term paradigm to denote a generally accepted set of assumptions and procedures that serve to define both subjects and methods of scientific inquiry[7]. When the assumptions and the procedures which serve them cannot answer anymore the aim for which they have been formulated or when the aim seems in need of reformulation as the answers are not adequate anymore, new assumptions and procedures have to be formulated which might answer the questions posed. Where Kuhn states that unanswerable questions lead to revolutionary science I would say that at any given moment in time a paradigm meets its paradoxical[8] frontiers. When the paradoxes accumulate there comes a moment when it seems logical to formulate a new paradigm, which inherently requires new practises[9]. The idea of paradigm shift cannot, however, be formulated too strictly. As knowledge and the data it is based upon are cumulative and assumptions and procedures follow each other consecutively there can be no sharp divide between one paradigm and another. It can sometimes be that some older assumptions are still valid and some procedures still workable within a new paradigm. Only with historic hindsight can one tell when one paradigm has been replaced totally by another. Furthermore a paradigm shift in itself might be an objective occurrence, but it can only be acknowledged as such when there is wide subjective cognizance of the new paradigm.

The problem for us is that we are probably in the middle of a paradigm shift and therefore will have problems defining what assumptions and procedures might be adequate in the future and in what way we can make the transition as smooth as possible, as we cannot imagine future use except in a probabilistic way, presuming we can at least transfer contemporary procedures to future technologies. It feels like falling into a science-fiction story or a fairy tale, where part of the environment (physically and/or mentally) reminds us of our normal or traditional environment, but where there are also many phenomena and modes of thinking, living and communication which are hard to interpret, because we have to imagine them at the vague frontiers of our intellect and emotion. Sometimes we do not

7 Kuhn phrases it as the “universally recognized scientific achievements that for a time provide model problems and solutions to a community of practitioners”. Kuhn, T.S. (1962). *The structure of scientific revolutions*. Chicago, University of Chicago Press, p. X.

8 Paradox: a person, thing or phenomenon at variance with normal ideas of what is probable, natural or possible. Hayward, Arthur L. and Sparkes, John J. (1982). *The concise English dictionary*.

9 Or as Lyotard phrased it: science comes more and more to the conclusion that her dynamism relies on the paradox, on incompatibility and indecisiveness. Peperstraten, Frans (2000). *Jean-François Lyotard*. In: Filosofen in deze tijd, p. 201.

understand them because our frame of reference does not allow for them or because we are afraid to let go of our traditional frame of thinking and imagining. New modes of communication, or in our case defining new kinds of queries and answers, can be especially dumbfounding as we do not yet fully understand the underlying syntax and semantics, nor what they should signify. And most of us are not aware yet that while looking for the golden apples, guarded by the Hesperides, Atlas is going to put the burden of the world on our shoulders. And it might take a long time before we have outwitted him and return the burden.

Kuhn was chided for his paradigm theory, especially the idea of incommensurability of subsequent paradigms, as he could only give a historical review of the phenomenon. This might be a valid criticism were it not for the fact that during his lifetime no total paradigm shift was observable, as is probably also applicable to our own situation[10]. I think his theory applies to our situation as the basis on which our assumptions seem to work (the data), the processing of them (software agents), and the way in which we have to manage the production flow (TGIS[11]), as well as the aims for which these procedures are tools are changing in such radical ways, that the future situation seems incomparable or incompatible with the present or traditional one. Though the assumptions and procedures of the different paradigms are incommensurable they will exist side by side for some time as some of the data within the former paradigm, especially DLOs, cannot be transformed into the new paradigm, as the data from which they are constructed cannot be extracted in separate entities with their accompanying attributes from them and thus cannot function as independent data within the new paradigm. The digitized maps need to be emulated[12]. But also certain procedures ought to have low-level processing and retrieval agents, which are emulations of procedures of the former paradigm.

One of the properties of a paradigm shift is that experimental or innovative assumptions and instruments will diverge, i.e. the gap between separate (groups of) collections will widen for some time as, in the first instance, each will define anew its goals and the ways in which to try to achieve them. As with the old paradigm

---

10 I think he should have applied his theory more to the totality of sciences or society as a whole than to a specific science. As a result paradigm shifts would take more time than ‘revolutionary science’ before new paradigms are accepted, as they have to permeate the sciences or society.

11 Temporal GIS. “Although the computer technology for spatio-temporal analysis exists, the GIS community must undergo a paradigm shift to fully appreciate TGIS benefits. It’s not just a matter of collecting time-based data within a GIS, but also developing a new way of thinking about time in a spatial sense”. Jim Castagneri (1998). Temporal GIS explores new dimensions in time. http://www.geoplace.com/gw/1998/0998/998tmp.asp [Accessed: 20 February 2001]

12 Rothenberg, Jeff (2000). *Using emulation to preserve digital documents*. The Hague, Koninklijke Bibliotheek, 69 p.

(Paris Principles and ISBD/MARC) it will take time to converge upon common goals and instruments. Some collections may have more trouble achieving these than others, depending on feelings of certainty about the goals to be achieved and resources demanded and being available.

In order not to lose hope or get lost we have to assess the available contemporary technology and as there is no other way try to look for ways to emulate, migrate or transform them to newer configurations. At the same time we have to try, however, to innovate technologies or pirate them from other science-fields in order to be able to answer new kind of queries. Unfortunately I'm not practically involved sufficiently in how databases and the relations and topologies of their data are constructed and how these are developing, to totally understand and describe the change. The only thing I can do for the moment is try to imagine how they will develop without going into particulars. On the other hand I can summarize the traditional retrieval mechanisms we use and see how additional and/or enhanced functionality can be developed, and comment on the possible adaptability to our field of new retrieval mechanisms from other fields. When speaking of traditional retrieval mechanisms I think of Dublin Core (DC), ISBD/AACR-MARC (IAM), FGDC/ISO (FI)[13] standards and the like; when talking about innovative retrieval mechanisms I think of tools like visual geographic interfaces, TextTilling, and Knowledge Discovery and Data Mining.

From DLOs to data

The traditional paradigm we are working within is that of identifying DLOs. What we have to try to cope with is that in the future instead of images constructed of several layers of data printed on top of each other within a single image we will primarily work with single data or datasets. This is a change with more impact on our work than we may imagine. Up till now we have been well versed in reading cartographic representations and translating their contents to our customers who might not be so well versed. We do not have to be cartographers because the image is constructed for us and we mainly have to evaluate whether the sources used in constructing it are up to par for the query posed. We have learned to read the patterns and relations in the single and combined features of the map. We have learned to access the image through its particulars, using its formal descriptive elements and interpreting its cartographic and mathematical properties and subject matter. For interrelations we use geographic and subject thesauri. And when we have a sophisticated retrieval system we can select and combine single features and

[13] It might be that the FI-standards are mechanisms, which are exemplary for the paradigm shift, while they are innovative in coding nuclear entities and their attributes or properties, and qualifying them.

properties from the descriptions of the items and as an answer to our query we get one or more descriptions of maps that contain the information we are looking for. When necessary we can aid our customer in interpreting the documents he/she gets as a result of the query that was posed.

This method, however, does not always gives the desired answer, and usually will only be an approximation to the query posed, as no single item or set of cartographic items fully answer the question(s) posed. This is because we are depending on the selection and interpretation of data by the cartographer who constructed the map. And in the construction from basic to generalised data the cartographer not only uses one to one mechanisms, but also adds or distracts to/from the resulting images in such a way that the image becomes more friendly, artful or usable for the user[14].

The array of products which are produced, use but a fraction of the amount of data available, as physical production of maps is limited. Some data is probably never used as nobody imagines that it can be used for the construction of maps, or because it is economically not viable to use it, or because simply nobody knows that it exists or can be functional for certain purposes.

In his work *Laocoon*[15] the German dramatist and critic G.E. Lessing (1729-1781) "warns poets against empirical descriptions of landscapes, because the consecutive nature of language breaks up into parts what is bodily coexistent. These parts are then difficult or impossible to put together again. What exists side by side becomes transformed into before and after by verbal description. Painting has not this kind of limitation. Nor has the map, a geographer would add."[16]. However, the map is an interpretation of only a small part of the data available to construct it. As soon as it is created it gives the overview which hopefully is an approximation of the reality we seek, meaning we can far more easily interpret the interrelations between its constituent parts.

From evaluating to creating

But the future is probably inversely different. "With an array of analysis tools [researches in future] will work more exhaustively with the … digital resources than they have so far been able to work with analogue ones. … Increasingly, research efforts to interpret these data are assisted by data visualization tools. In

---

14 The cartographer can even differ from reality when he puts the several layers of a map (e.g. topography, orography, text, etc.) together to make the image more legible or artful.

15 Lessing, Gotthold Ephraim (1766). *Laokoon oder ueber die Grenzen der Mahlerey und Poesie*.

16 Hägerstrand, Torsten (1995). *Landscape as overlapping neighbourhoods*. In Benko, Georges B., and Strohmayer, Ulf. *Geography, history and social sciences.* Dordrecht, Kluwer, p. 86.

addition to data mining and visualization tools, future researchers will be aided by intelligent agents that explore the … [data] looking for information that meets certain user-specified criteria and refine their searches as they accumulate data and knowledge. Digital archives [or repositories] combined with new technologies will liberalize scholarship. They will enable simultaneous access to a range of sources (both local and distant) and facilitate the use of research methods not possible with conventionally printed or hand written records."[17]

Though many of our colleagues understand the words in the premise Andrew Tatham laid down during the LIBER-conference in 1994[18], few of us can yet act it out in reality or understand its full impact.

In the data-age we, or software which acts in lieu of us[19], have to interrogate our customer as to what kind of features, topology, relations, patterns, etc. he/she wants to view in the image to be created. The easiest possibly is to select the area for which the image has to function, together with its mathematical properties, like scale, projection, geographic grid[20], etc. Then we have to try to fill in that empty space with point-, line- and polygon-features with their attributes from one or more datasets available to us. And we must keep in mind that some might be stacked one on top of the other as datasets might come in many levels of functionality[21].

Now we have to try to think what the image we want to create might look like, then retrieve the data with which this image can be constructed, and then use the software that actually constructs this image. This means we have at least to be part-time cartographers to effectuate this process, and keep the critical words of Lessing echoing in our minds. The image to be created should be first formed in our mind's eyes before we select its constituent data to build it, in order that the

---

17 Ross, Seamus (2000). *Changing trains at Wigan: digital preservation and the future of scholarship*. London, National Preservation Office, pp. 11-12.

18 "*We shall no longer provide the users with someone else's selection and presentation of data, but with the data itself and with the means by which the user can make their own selection and presentation of this data to inform or to mould their own or other people's image of the world*". TATHAM, Andrew (1995). *Can the map curator adapt*? In: The Liber quarterly, vol. 5/1995, no. 3, pp. 330-336. [http://www.kb.nl/infolev/liber/articles/2tatham.htm]

19 It is questionable whether idiolect can be coded into software, in order to retain part of our subjective frame of reference, unless this software has the innate ability to learn from interaction with the customer. If so, for regular customers, their own query-profile might evolve, which includes part of their own idiolect.

20 Most of this is 'spatial reference information'. E.g. Phillips, Hugh (2000). *'Colorized' Content Standards for Digital Geospatial Metadata, FGDC-STD-001-1998*. [http://badger.state.wi.us/agencies/wlib/sco/meta/colorstd/hier_basiccsdgm.html#Spatial_Reference_Information. Accessed: 2 January 2001]

21 E.g. *TOP10vector*, the 1:10,000 digital topographic map of The Netherlands, consists of 69 layers of information, which might be retrieved per layer or in a combination of layers.

cartographic overview we want to create is not marred by the fact that we fall into the trap of what is before and what is after without knowing what is the whole.

But it also means a radical change in accessing the data we have under our control. Descriptions like DC and IAM are inadequate as they are developed for identifying DLOs. Descriptions like FI are more adequate as they are developed in order to be able to evaluate the inherent quality of digital data or datasets. But they are first-generation standards, which can be used for evaluating and selecting single datasets. Are they also functional for knowledge discovery and data mining, i.e. finding the relevant data-elements or collections of data-elements from disparate datasets, which will be functional in the map-image our customers are requesting[22]?

The new methods of processing depend for a large part on newly formulated queries and the belief or acceptance that such queries should be answerable with the data or datasets available. This means that new retrieval technology will develop experimentally and that many developments will turn out to be dead ends. To prevent too many dead ends we have to look at technologies used in those disciplines which work with vast amounts of data, like astronomy or remote sensing, census processing or the actuary field. In these disciplines testing tools try to make out patterns, relations and topology.
Geographic data growth is more than exponential. When we call a one-million map collection large by present day standards in Terabyte measures it might be called not overly large, and in future it might be even called small, also taking into account that a lot of the DLOs are topographical maps. The problem is that our tools with which we manage future data-collections must be more automated and partly independent as the human mind can but cope with a limited amount of data and produce from this a limited set of DLOs. In order to be able to manage vast quantities in the future we have to be prepared to become more and more meta-managers, of the data as well as of the metadata.

And we cannot stick to maps alone, though these are historically the background to our profession. We must be willing to widen our perspective to include pure geographic data, thus bringing geography and cartography back to their

[22] N.B. data mining is data-driven, whereas by contrast, traditional statistical, reporting and query tools are user-driven. To be a true knowledge discovery method, a data mining tool should unearth information *automatically*, and extract these in such a way that they can be put to use in areas such as decision support, prediction, forecasting, and estimation. I will not investigate in this article the impact of this technology as this is still too far away from our everyday practices.

dichotomous nature[23]. But also we might sift auxiliary data from fields[24] which are not geographic in nature but which collect or use much locational data.

Within this frame of mind we have to look at the same time at combining or recombining disparate data of different value and functionality or even better, imagining the data and datasets as an amalgamated whole, in order to come to new insight or answers. However, we have to strive within our competence, while at the same time trying to imagine new, yet unknown, goals in our minds. Some of these processes might be autonomous as TGIS together with data mining can analyse data and their combinatorial values. One of the results might still be a cartographic map in order that the human mind can get an overview and process the data visualised within the image.

Assessment of traditional access-tools

When the documentation community became aware of the fact that electronic data or datasets needed a different approach of documentation then traditional analogue materials a vast amount of initiatives and evaluations took place in the 1990s. Since the early 1990s producer- and user-communities of elaborate digital datasets (DEM's, TIGER-files, remote-sensing images, etc.) started to develop sophisticated standards to document these.

In libraries bibliographic processes have been aimed at the user-tasks: 'find', 'identify', 'select', and 'obtain'. Other metadata-initiatives have added amongst others 'qualify' and 'manage' or 'housekeeping'.

It is understood that information retrieval is measured in terms of recall and precision. If a lot of relevant information is missing, there is poor recall. If we get flooded by a lot of irrelevant information, there is poor precision. This pronouncement is adequate for discerning differences between the relevant metadata systems, which will be described below.

To foster understanding I have created a diagram in which the metadata-systems have been divided into 4 groups. Added is a 5th group which contains innovative tools, which will be treated separately.

---

23 And incidentally give more value to the "G" in the name of the Geography & Map Library Section.

24 Like biochemistry, biology, botany, engineering, history, medicine, physics, psychiatry, zoology, etc. See: Millea, Nick (1998). *Delivering digital data into the library: the DIGIMAP project and its impact on the map room - the Bodleian Library experience*. In: The Liber quarterly, vol. 9/1998, no. 2, pp. 189-200, table 3 [http://www.kb.nl/infolev/liber/articles/millea03-11.html. Accessed: 2 January 2001];
Morris, Barbara et al. (2001). *EDINA Digimap: new developments in the Internet mapping and data service for the UK Higher Education community*. In: Liber quarterly, vol. 10/2000, no. 4, Table 1 and 2, pp. 446-448.

Quality	**Simple**	=========>	=========>	**Rich**	**Innovative**
Quality Level	**Band I**	**Band II**	**Band III**	**Band IV**	**Band V**
Diffusion	=========> +++++++++>	=========> +++++++++>	<========= ==========> ++++++++++>	<========= +++++++++>	<+++++++++
Availability	Internet	Internet	Internet / Intranet	Internet / Intranet	Intranet
Purpose	Location	Selection	Evaluation	Analysis	Intelligent analysis
Unit of information	Individual digital information object	Logical set of digital objects; no links between documents	Publication; links between whole and parts	Databases with links between whole and parts on all levels	Databases with geometric links between whole and parts on all levels
Standards	Proprietary	Emerging standards	Generic standards used in information world	Standards used in specialist subject domains	Intelligent software-modules
Form of record	**Proprietary simple records**	**Dublin Core**	**AACR2, ISBD**	**FGDC, CEN, ISO, ANZLIC**	Not applicable
Format	Unstructured	Attribute value pairs **Dublin Core**	Subfields, qualifiers **MARC USMARC UKMARC UNIMARC MARC21**	Highly structured mark up **FGDC DTD ANZMETA**	Deriving from Knowledge Discovery and Data Mining
Conservation			**Nedlib**	**Nedlib**	Not applicable
Input	Robot generated	Robot plus manual input	Manually input	High level of manual input	Intelligent expert system
Conversion		DC/MARC/GILS	MARC2DC MARC2FGDC	FGDC2MARC	Not applicable
Protocol	http with CGI form interface	directory service protocols (whois++) with query routing (Common Indexing Protocol)	Z39.50	Z39.50 (in future with collection navigation)	Collection navigation

Diagram of access- and retrieval systems

Band I[25]

These are simple records, created mainly by robot search engines by general search-and-auto-describe missions. As they do not discriminate between the functions of the words they harvest the resulting indexes show a lack of precision for those seeking qualified information. Unaware web surfers who start out to explore the available information in a random way mainly use them.

Band II

On the instigation of OCLC and NCSA workshops have been held since 1995 to try to find a modus in which these metadata can be formulated. The first workshop held in Dublin, Ohio (U.S.A.) found consensus on a set of elements called since then the **Dublin Core** (DC).

25 For a more extensive treatment and developments within Bands I-IV see: Smits, Jan (2001). *Metadata and standards, confusion or convergence?* In: ???

Dublin Core records are specifically created for resource discovery and as such description requirements are generally less precise than for traditional bibliographic records. This, in order that also non-professional documentalists can generate them. Soon the user community (among others the library field) found, however, that more precision could be needed to be better able to query DC-databases. Precision is sought in qualifiers, which can be added to metadata-tags. And when possible using controlled systems[26]. A problem to be solved is that most controlled systems or vocabularies (like Dewey Decimal Classification 21, Universal Decimal Classification, Library of Congress Subject Headings, etc.) are not freely available on the Internet, in which case those users who have disposal of these can only use their quality. Use of these controlled systems would enhance interoperability with other metadata-systems, like those in Bands III and IV.

Libraries use DC-metadata in an explorative way but also should embed them in the electronic resources they create themselves and make available through the Internet. Too much metadata is created in a post-coordinate process. To create more uniformity of DC-records some organisations provide more or less sophisticated templates[27] to enhance the value of the metadata.

As the Dublin Core is maturing a little the user-community is discussing whether the 15 Dublin Core elements might be more coherently expressed if they are related to an underlying logical model, which treats information resources as having logical states (an abstract work or a physical item, for example) that have relationships to each other and to other resources[28]. The inherent simplicity of DC-cataloguing poses quite some problems to map librarians, mainly due to the fact that there is no consensus yet as to what is 'best practice' and 'minimal adequate' DC-cataloguing. Though the syntax of DC is developing quite well there is a lack of semantics, as there is no rulebook like the ISBD.

---

[26] E.g. <DC:coverage CONTENT="(SCHEME=INTERNATIONAL PRIME MERIDIAN, GMT)"> W123°20´- W121°09´/ N49°38´- N48°55´</DC:coverage>

[27] E.g. Koch, Traugott and Borell, Mattias (1998). *Dublin Core Metadata Template* <http://www.lub.lu.se/cgi-bin/nmdc.pl> [Accessed: 25 January 2001], part of the 'Nordic metadata project'. These and other tools can be found on the website 'Metadata Related Tools'. DCMI <http://purl.org/dc/tools/index.htm> [Accessed: 25 January 2001].

[28] Weibel, Stuart (1999). The State of the Dublin Core Metadata Initiative April 1999. *D-Lib Magazine*, Volume 5 Number 4. CNRI < http://www.dlib.org/dlib/april99/04weibel.html> [Accessed: 25 January 2001]

Band III

In 1961 the IFLA adopted the Paris Principles[29] as basis for an international approach to "headings" and "entry words". They were however solely meant for catalogues of printed books. In 1967 the Anglo-American cataloguing rules (AACR) were published, followed by an interpretation manual for mapcurators in 1982[30]. The convergence of cataloguing codes started in 1971 when the IFLA published the first of the ISBD's for monographs and serials soon to be followed in 1977 by the ISBD(CM) for cartographic materials. Almost all ISBD's were revised in the 1980s and 90s (ISBD(CM) in 1987) and the last to appear in 1997 was the ISBD(ER) for electronic resources. This latter created havoc, because it means that most ISBD's have to be revised again to align them with the new insights concerning the description of electronic materials.

But there is more to it. As a result of changes in the expression of documents and their transport media IFLA started to evaluate the whole configuration of Paris Principles and ISBD's[31]. This study evaluated the value of the fields and elements, using a user-oriented conceptual model, where the elements are described as entities, which have attributes (the attributes serve as the means by which users formulate queries and interpret responses when seeking information about a particular entity) and relations to other entities. When they are related to user-services their function is still restricted to find, identify, select and obtain.

The ISBD(CM) under revision, of which a third draft was circulated in May 2000[32], tries to amalgamate alterations to the ISBD(CM) and those parts of ISBD(ER) which seem appropriate. The draft is largely adapted to the recent issue of CCQ[33]. For the moment there is a choice to extend area 3 (the material specific area) with:

- Statement of accuracy [for scale] (optional)
- Designation and structure of electronic resource
    Structure of resource (optional)

---

29 IFLA Committee on Cataloguing (1971). *Statement of principles, adopted at the International Conference on Cataloguing Principles, Paris, October, 1961*. Annotated edition with commentary and examples by Eva Verona. IFLA.

30 Stibbe, Hugo P. (1982). *Cartographic materials : a manual of interpretation for AACR2*. ALA.

31 IFLA Study Group on the Functional Requirements for Bibliographic Records (1998). *Functional requirements for Bibliographic Records* (UBCIM publications - new series, Vol. 19). IFLA. Also available through <http://www.ifla.org/VII/s13/frbr/frbr.pdf> [Accessed: 25 January 2001]

32 Bäärnhielm, Göran (ed.) (2000). *ISBD(CM): international standard bibliographic description for cartographic materials*. Revision including Electronic Resources. Third draft. (Not published).

33 Andrews, Paige G. and Larsgaard, Mary Lynette (1999). *Cataloging & Classification Quarterly*, Vol. 27. Haworth. Also published as: *Maps and related cartographic materials: cataloging, classification, and bibliographic control*. New York, Howarth, 487 p.

Number of files, records, bytes (optional)

- Digital graphic representation method (optional)

  Object type (optional)

  File format (optional)

  Object count (optional)
- Geospatial reference data (optional)
- Designation of electronic resource (when applicable)
- Characteristics (i.e. structure) of electronic resource (optional)
- Digital graphic representation data (optional)

Further there are large changes/additions in/to the 'edition statement', and 'notes on system requirements'. Though the latter seem to be mandatory in all ISBD's I wonder if they will not serve better under the 'manage' part of metadata. Especially after a certain time when most electronic resources need to be converted, emulated, migrated or otherwise changed these data serve mostly internal library functions and not user-needs.

AACR and the ISBD are cataloguing codes, but before the user can search the bibliographic databases and read the descriptions they have to be electronically processed. For this purpose MARC (MAchine Readable Catalogue)-formats are developed. MARC is an implementation of the international standard *Information and documentation - Format for information exchange* (ISO 2709-1996)[34].

To alleviate problems with existing MARCs the Library of Congress in consultation with various user communities develops MARC21. This format "is an integrated format defined for the identification and description of different forms of bibliographic material. MARC21 specifications are defined for books, serials, computer files, maps, music, visual materials, and mixed material. With the full integration of the previously discrete bibliographic formats, consistent definition and usage are maintained for different forms of material"[35]. The format attempts compatibility with e.g. UKMARC and UNIMARC. After consultation with the UK library and information community the BL Executive Committee has decided to adopt the MARC21 format to replace UKMARC in time. One of the reasons for

[34] The British Library (1999). *MARC home page*. BL <http://www.bl.uk/information/marc.html> [Accessed: 25 January 2001]

[35] MARBI (1996). *The MARC 21 Formats: Background and Principles*. LoC <http://www.loc.gov/marc/96principl.html> [Accessed: 25 January 2001]

this decision was that "MARC21 represents a more effective route to the eventual adoption in the future of non-MARC metadata standards …"[36].

With the upsurge of Internet applications the Library of Congress started in 1995 the project 'MARC DTD'[37] to enable the conversion of MARC-data to an SGML-environment and vice versa. With this DTD MARC-data can be converted to the Internet, but also organisations can create MARC-like metadata in the Internet and convert it to MARC-databases.

Band IV

Digital spatial databases have been created from the late 1970's onwards and reach nowadays the phase where there is countrywide coverage (on municipal, provincial and state level) and through GIS they can be integrated with other databases. Initially they were a continuation of existing analogue processes, but for the fact that they are more in vector-format and are built up of layers of information, which can be manipulated independently from each other or in concert with each other. To extend operability a lot of producers are digitising their existing analogue data, usually in raster-format. Seeing the benefit of promotion of and the need for a higher return on operating costs producers started to think of ways to make these data more accessible. As the economic stakes are higher then before they sought to create a descriptive system, which incorporates not only data usually associated with ISBD's but also data, which could help users to evaluate and analyse the fitness of use and quality of the digital spatial data offered.
Examples of these standards are ANZLIC[38] (Australia & New Zealand), FGDC[39] (USA), CEN[40] (Europe) and ISO[41] (global). The latter two probably will come formally into existence in 2001. The ICA Spatial Data Standards Commission will

---

36 The British Library Executive Committee (2001). *MARC harmonisation* : British Library to adopt the MARC21 format : Survey results. http://www.bl.uk/services/bsds/nbs/marc/result1.html [Accessed: 29 January 2001].

37 Network Development and MARC Standard Office (1999). *MARC DTD's, document type definitions, background and development*. LoC <http://lcweb.loc.gov/marc/marcdtd/marcdtdback.html> [Accessed: 25 January 2001]

38 ANZLIC (1998). *Core Metadata Elements for Land and Geographic Directories in Australia and New Zealand*. ANZLIC <http://www.anzlic.org.au/metaelem.htm> [Accessed: 24 April 2000]

39 FGDC (1998). *Content Standard for Digital Geospatial Metadata* (FGDC-STD-001-1998). FGDC <http://www.fgdc.gov/metadata/contstan.html> [Accessed: 25 January 2001]

40 CEN/TC287 (1999). *The Geographic Information European Prestandards and CEN Reports*. AFNOR <http://forum.afnor.fr/afnor/WORK/AFNOR/GPN2/Z13C/PUBLIC/WEB/ENGLISH/pren.htm> [Accessed: 25 January 2001]. To view the full pre-standard you need to contact a CEN-member.

41 ISO (2001). *ISO/TC 211 Programme of Work*. ISO < http://www.statkart.no/isotc211/pow.htm> [Accessed: 25 January 2001]. To get access to the documents and the document catalogue, you need user-id and password. Please contact your national body to get this information.

publish in 2001/2 a book, which will compare all available national and international metadata standards and –systems, alike to their book on transfer standards[42].
As a lot of data is needed for the evaluation- and analysis-processes the records can run in the hundreds of elements and be anywhere between a few thousand and tens of thousands bytes. As such they can be also used as tools to administrate the datasets they describe.
Though one of the complaints concerning records created with these standards could be that they need a high amount of manual input I guess that better programming of the metadata-module within spatial datasets could help automated processing of numerical values (e.g. object-count) and other automatically generated data (e.g. geometric base values) in the metadata-records.

Towards meta-information systems?

In band IV-descriptions it is already possible to add metadata concerning the administration of the metadata. The metadata itself can provide information to assist in developing migration/emulation processes and to check whether these processes have functioned adequately. Metadata can help in ensuring the level of integrity of the data after necessary manipulations to preserve them for future use. They help to answer the question whether the data have the same information value and quality as before the manipulations took place and at the same time whether unavoidable losses can be qualified[43].

This drive is also detectable in the NEDLIB[44] (Networked European Deposit Library)-project, funded partially by the European Commission's Telematics for Libraries Programme. One of the reports[45] concerns itself with metadata for long term preservation of electronic publications, based on the OAIS information model[46]. The different kinds of metadata described in this report are:

---

[42] Moellering, Harold (ed.) (1997). Spatial database transfer standard 2: characteristics for assessing standards and full descriptions of the national and international standards in the world. Pergamon.

[43] Smits, Jan (1999). *Metadata: an introduction*. In: Andrew, Paige G. and Larsgaard, Mary Lynette. *Maps and related cartographic materials: cataloging, classification and bibliographic control*, p. 313.

[44] While supplies last all reports (up till now some 7 reports have been published) can be requested free of charge from the Koninklijke Bibliotheek by filling in the Nedlib-orderform at www.kb.nl/nedlib/

[45] Lupovici, Catherine [and] Masanès, Julien (2000). *Metadata for the long term preservation of electronic publications*. Koninklijke Bibliotheek. (Nedlib report series ; 2).

[46] CCSDS 650 0-R-1: *Reference model for an Open Archival Information System(OAIS)*. Red Book. Issue 1. May 1999. http://ssdoo.gsfc.nasa.gov.nost/isoas/ref_model.html [Accessed: 5 February 2001]

1. Descriptive metadata
2. Administrative metadata recorded for the deposit system management purposes
3. Metadata for preservation

The descriptive metadata can be the same as generated in band II-IV-descriptions, the administrative metadata can be the same as that generated in Band IV-descriptions. The metadata for preservation are a new subset and are used to monitor the actions needed for the long-term preservation of the content information. In the Nedlib report the following elements are given for a "core preservation metadata set"[47]:

- Specific hardware requirements
- Specific microprocessor requirements
- Specific multimedia requirements
- Specific peripheral requirements
- Operating system
- Interpreter and compiler
- Object format
- Application
- Reference information
- Assigned identifier
- URL
- Fixity information
- Checksum
- Change history
- Main metadata concerned
- Tool
- Reverse
- Other metadata concerned

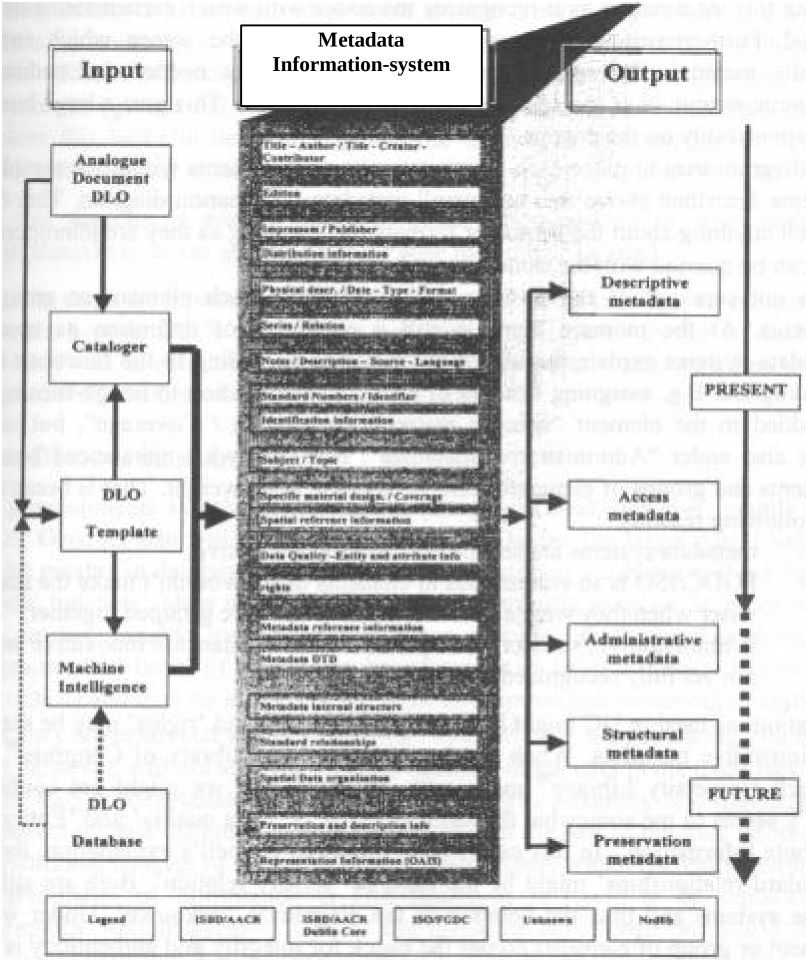
When we take a closer look at some of these elements we see that they have the same function as the data concerning electronic publications in the ISBD(ER), except that in the latter they are relegated to the note-area and are summarized within one element. The Nedlib report is ahead of a possible extensive expansion of the formats used for bibliographic library systems, which eventually might evolve into all-encompassing meta-information systems.

Not all of the necessary metadata needs to be manually processed. When one imagines the repository as a totally integrated system of DLOs and metadata with all relevant links available than an automated processing system can retrieve all the

[47] Lupovici, Catherine. Ebenda, pp. 17-21.

necessary metadata from the actions implemented for its preservation in the repository and add these to the relevant metadata-record(s). Instead of relying entirely on the prescience of the cataloguer, a digital library might allow direct search of the contents of an information object, particularly for domain-specific information or minor detail that a cataloguer might normally ignore[48].

[48] Goodchild, Michael F. (1995). *Alexandria Digital Library. Report on a workshop on metadata held in Santa Barbara, California, November 8, 1995*. http://www.alexandria.ucsb.edu/publicdocuments/ metadata/metadata_ws.html [Accessed: 15 February 2001].



Amalgamated diagram for the main metadata-concepts

When we take a sharper look at the metadata-systems we can see a lot of similarity, but also a lot of divergence. Not so much in the content of the metadata as well in the way they function at the input- or the output level.

At the Input side one must image that at the moment physical information resources and DLOs are offered to be accessed. Most of these are still manually processed by cataloguers. Some of the data the cataloguer extracts from a template

which is included in DLOs. But one can also image that a computer-program extract this information as it recognizes the codes with which certain elements are tagged. Furthermore scripts may be running behind the screen which extract specific metadata for specific purposes which are not needed for traditional catalogue-output, as is the case e.g. with the Nedlib data. This puts a large burden of responsibility on the creator.
The diagram tries to place each element or group of elements within the metadata-systems described above into an overall metadata-information diagram. This does not tell anything about the format or formats used, as long as they are interoperable and can be queried with the same protocols.
I am not sure yet of the position and function of each element or group of elements. At the moment there is still a confusion of definition as specific metadata-systems explain the use of the metadata according to the functions they get assigned. E.g. assigning headers of technical information to image-files might be added to the element "specific material designation / Coverage", but might come also under "Administrative metadata". It is somewhat unbalanced because elements and groups of elements seem to have the same weight. That is because of the following reasons:

- metadata systems are still predominantly descriptive;
- FGDC/ISO is so voluminous in elements that it wouldn't make the reader wiser when they were all included and so they are grouped together;
- Administrative, structural and preservation metadata are innovative and not yet fully recognized.

'Relation' as used in DC might be structural metadata; and 'rights' may be seen as administrative metadata. When we consider what the Library of Congress[49] and Cornell University Library[50] mean with structural data we might get confused. LoC's seems to me somewhat the same as FGDC's 'Data quality' and 'Entity and attribute information'. In this case I have preferred Cornell's explanation, though 'standard relationships' might be the same as 'series / relation'. Both are still in-house systems and thus the notation in the diagram is 'unknown'. Under which element or group of elements comes the check for integrity and authenticity is also uncertain.

---

49 LoC (1999). Structural Metadata Dictionary for LC Digitized Material. Version 1.03. http://memory.loc.gov/ammem/techdocs/repository/structmeta.html [Accessed: 21 February 2001]

50 Cornell University Library (2001). Moving theory into practise : Digital imaging tutorial. Chapter 5: Metadata http://www.library.cornell.edu/preservation/tutorial/metadata/metadata-01.html [Accessed 21 February 2001]

Seen from the output side it might be necessary to extract metadata from different formats, as this is functionally dependent. Should crosswalks between the diverse formats be necessary to effectuate certain output this should be done behind the screen without bothering the user. The five output systems are put tentatively opposite the (groups of) elements from which they extract their metadata. At the moment this looks to be biased against the innovative systems, but in future I think, when seen the amount of metadata used, the opposite will materialize.

But this diagram is mainly meant to foster discussion as to what metadata might/should/can cover and for what purpose they might be used.

In a later stage the input-side will change radically. In the case that all information resources become DLOs or databases metadata will be inherently part of them and obviously will be extracted mainly by computer programs. Catalogers then will be assigned the functions of checking and validating.

Again DLOs vs databases

The developments sketched in the above paragraphs deal, however, mainly with DLOs. Geodata also will be partly created as DLOs, but the larger part, I believe, will be created in databases, where the entities and their attributes will not have a specific function unless they are generated for a specific purpose. In other words they are but variable instantiations within a specific timeframe. The generated images or other forms of output may be referred to as fixed document like objects (see note 4) and may be in need of the same description and preservation metadata when they are preserved within a repository system[51]. Such a generating process in itself, however, will not alter the inherent qualities of the data in the databases.
"In the long term metadata … will evolve to support a hierarchy of different levels of abstraction, and different degrees of expertise on the part of the user. Moreover, the traditional concept of the map sheet or image scene, which currently dominates the granularity of information is GIS-databases, is likely to be replaced by geographical seamlessness"[52]. One of the problems which will produce a lot of discussion in the near future is the *granularity*[53] of information to be accessed, i.e.

---

[51] I believe that examples of these generated fixed DLOs should be preserved in order that future users can obtain knowledge of part of the historical context during the lifetime of the databases.

[52] Goodchild, Michael F. (1995). Ebenda.

[53] A meta-data entity that is associated with the lowest level of granularity of information available to InfoHarness is called the information unit (IU). The IU may be associated with a file (e.g., a man page), a portion of a file (e.g., a C function or a database table), a set of files (e.g., a collection of related bitmaps), or any request for the retrieval of data from an external source (e.g., a database query).

what is the detail of the unit of information which should be considered to merit specific descriptive attention. “In the digital world, granularity of information may take on an entirely new and unfamiliar forms that are no longer linked to the granularity established by the author or publishers”[54].

Producers of geographical databases will find a descriptive format to describe not only their contents[55], but what is more important, to describe the potential uses to which the data in their databases can be put to. What really is lacking are search strategies for the data to mold one’s image of a specific or combination of (a) geospatial phenomeno(a)n or how to find the spatial data which can help us with decision support, prediction, forecasting and estimation of a specific spatial phenomenon or combinations of spatial phenomena.
For such processes metadata are inadequate as we are interested then in the inherent qualities of spatial data and their attributes, topology to or combination with other spatial data. This also independent of the geometric and geographic constraints a producer might have put to demarcate its database. And then we haven’t thought yet about the locational data, which may guide us to non-geospatial resources, which can help us to put a possible geospatial image into an overall context.
To imagine how innovative concepts interact with our traditional view of accessing data I have added Band V to the diagram. But a thick line from the other four bands separates this band. It is questionable whether it can be added to the diagram as the functioning which can be described therein might be incommensurate with the other bands and might belong to another paradigm. But for the sake of proposing innovative technologies I have added it anyway.
The next paragraphs will deal with the possible innovative technologies which might help us to find the ‘gems of information’ we are looking for.

Making locational data available

Locational information can be found in many more resources than only cartographic materials or geospatial databases. When we look at documents and databases concerning statistics, history, biology, ecology, travel, etc., many of

---

Leon Shklar, Leon et al. (1994). *“InfoHarness" Information Integration Platform*. http://www.ncsa.uiuc.edu/SDG/IT94/Proceedings/Searching/shklar/shklar.html

54 Goodchild, Michael F. Ebenda. One can imagine that automated modules tag every distinct unit of information within a database with metadata which is taken from the structuring principles of this database, e.g. geometric, representational and/or qualitative attributes. When we come to knowledge discovery and datamining this might well be necessary, but it does not need manual input!

55 For this purpose they can use the FI-standards.

these are concerned with a certain location on earth. In traditional library-catalogues these are mainly accessed by subject matter, but not according to geographical area. It is my opinion that a researcher is always searching for a body of interrelated information, which will provide knowledge for the subject he/she is working on. And we would like geographic information to be studied in context, which means textual and statistical resources should be available at the same time.
Of course we can use the traditional geographical- and subject dictionaries and thesauri for this purpose, but why not use the unique mathematical properties of maps in an age where visual information and the possibility of aggregating information in graphical representations are becoming more and more predominant. When we can translate subject thesauri for locations (like Amsterdam, The Netherlands, U.S.A., etc.) through a conversion table to bounding-box or point co-ordinates the same interface might be used for directing users to these non-map resources.

The problem is best described in a paper by Ray R. Larson[56], which discusses geographic information retrieval (GIR) within the Alexandria Digital Library project for both map-like spatial objects as well as georeferenced material. “As with traditional print libraries, […] information can be indexed and retrieved in a variety of ways, ranging from purely descriptive cataloguing of items in the database and topical analysis of content, to more specialized methods of classification and description that exploit the characteristics of digital information”. It not only describes the background but also discusses some tools for automated geo-referencing of text.
Larson describes and summarises the GIPSY-model. GIPSY, The Geo-referenced Information Processing System, was developed as a new model of automatic geographic indexing for text documents. In the GIPSY model, words and phrases containing geographic place names or geographic characteristics are extracted from documents and used to provide evidence for probabilistic functions using elementary spatial reasoning and statistical methods to approximate the co-ordinates of the location being referenced in the text. The actual "index terms" assigned to a document are a set of co-ordinate polygons that describe an area on the Earth's surface in a standard geographical projection system. Woodruff and Plaunt describe the GIPSY method for automatic geo-referencing in detail[57]. Later this method was evaluated, together with the POSTGRES method and the

[56] Larson, R. Ray (1996). *Geographic information retrieval and spatial browsing*. In: GIS and Libraries: Patrons, Maps and Spatial Information (ed. by Linda Smith and Myke Gluck), pp. 81-124 http://sherlock.berkeley.edu/geo_ir/PART1.html (preprint) [Accessed: 27 September 2000].
[57] Woodruff, A.G. & C. Plaunt (1994). *GIPSY: Geo-referenced Information Processing System*. In: Journal of the American Society for Information Science, 45, pp. 645-655.

TextTilling method in a paper for the Sequoia 2000 project[58] in the framework of the Global Change programme.
One of the next big projects within ADL is to take all the catalog records from the UCSB Library's online catalog that have geographic subject headings, then apply coordinates to each record, put the records in the ADL Catalog, and see if the resulting searches are as helpful to the users[59].

On the Internet several (proto)types of Geographic Information browser, which can perform this kind of function we are looking at, are available, e.g.

- The Environmental Resources Information Network [ERIN] Unit of Australia has developed a generic WWW map interface using a collection of simple map images and a standard lookup table to provide visual interactive access to geographically related information[60].

- MERI (Meadowlands Environmental Research Institute, New Jersey, U.S.A.) built a WWW-based interface to the database, integrating web server, database server and GIS server technologies. Through a map interface, a user can obtain a list of documents that studied a particular area. Conversely, through a text interface, a user can obtain a list of documents that report on, for example, a land use or cover type or a specific water body, with sampling/analysis locations displayed on a map[61].

An atlas as a geographical interface

As an example I would like to sketch a project that was drafted in co-operation with some university departments of cartography in The Netherlands. The idea is to have maps function not only as geographical information sources, but research whether it is possible to have the same maps function as visual interface for metadata-databases. For this project we looked at three possible databases, which could function together within this frame.

---

58 Larson, Ray R. et al. (1995?). *The Sequoia 2000 Electronic Repository*. In: Digital Technical Journal. 7(1995)3, pp. 50-65. http://www.digital.com/DTJJ04/DTJJ04SC.TXT (preprint) [Accessed: 28 September 2000].

59 E-mail communication by Mary Larsgaard, Map and Imagery Laboratory, Davidson Library, University of California, Santa Barbara (13 February 2001).

60 Crossley, David & Tony Boston (1995). *A generic map interface to query geographic information using the World Wide Web*. http://www.csu.edu.au/special/conference/apwww95/papers95/dcrossle/dcrossle.html [Accessed: 24 September 2000].

61 Barrett, Kirk R. , Richard Holowczak & Francisco J. Artigas (1999). *A database of environmental documents about an Urban Estuary, with a WWW-based, geographic interface*. http://www.awra.org/proceedings/www99/w21/ [Accessed: 27 September 2000].

1. NCGI[62]: This metadata-database of the National Clearinghouse Geo-Information contains presently more than 1,500 descriptions of digital spatial datasets with visual examples of some 17 producers of geospatial data. The datasets range from several 100s Mb to 10s of Gb.
2. AvN: The second analogue edition of the Scientific Atlas of The Netherlands (AvN, published 1984-1990) contained some 1,000 maps, which give a comprehensive view of the socio-economic, physical and ecological situation of The Netherlands. These maps are now digitised[63], but give not up to date information.
3. There is a database of some 30-40,000 descriptions of cartographic documents concerning The Netherlands within the framework of the Dutch Depositary collection in the Royal Library. The Royal Library phased out the CCK-system[64] in 1999 and will convert these descriptions between 2001-2002 to the PICA[65]-database.

The NCGI concerns mainly large-scale datasets with a high economic value. The owners of these datasets, which are usually (semi)governmental or academic bodies, are usually forced to recoup part or most of the costs involved in producing these data. At the same time these producers aggregate these data into the middle- and small-scale data necessary to create the maps used in the Scientific Atlas of The Netherlands. In the age of analogue products they did this service for free. But in the digital age it is hard to get the same services from them. This means we must find a modus to entice them in such a way that they are willing to create and provide these data for free. We think we can reach this goal by offering the maps of the Scientific Atlas of The Netherlands as a visual geographical interface for the NCGI.

The geographical interface would provide several search strategies:
1. When clicking on a map of The Netherlands one can search the underlying metadata-databases for datasets that also cover the whole of The Netherlands.
2. When making a cut-out of a part of The Netherlands one can search the underlying metadata-databases for datasets that also cover that part of The Netherlands.
3. When one zooms in the interface will query the underlying metadata-databases for datasets, which cover that specific part of The Netherlands.

---

[62] NCGI (2000). Geo-gids. http://www.ncgi.nl/Profiel/profiel.html [Accessed: 24 September 2000].

[63] Disciplinegroep Kartografie (2000). *Atlas van Nederland*. http://avn.geog.uu.nl/ [Accessed: 24 September 2000]

[64] VELDEN, G.J.K.M., P.J.M. Douma and J.G. Zandstra (1990): *CCK : making cartographic materials accessible*. In: The LIBER Quarterly 2(1992)2, pp. 192-208. http://www.kb.nl/infolev/liber/articles/cck.htm [Accessed: 24 September 2000].

[65] PICA (2000). *About Pica*. http://www.pica.nl/en/about/ [Accessed: 24 September 2000].

The queries will be effectuated by mainly using bounding-box co-ordinates. There must be, however, also possibilities for searching on point locations, within a radius of a point, a bounding polygon, etc., and when the computer can read the index-map also on named area, administrative sub-division or locality description, when necessary using a thesaurus. But it must be also possible to use traditional geographic and thematic dictionaries to come to the same result.

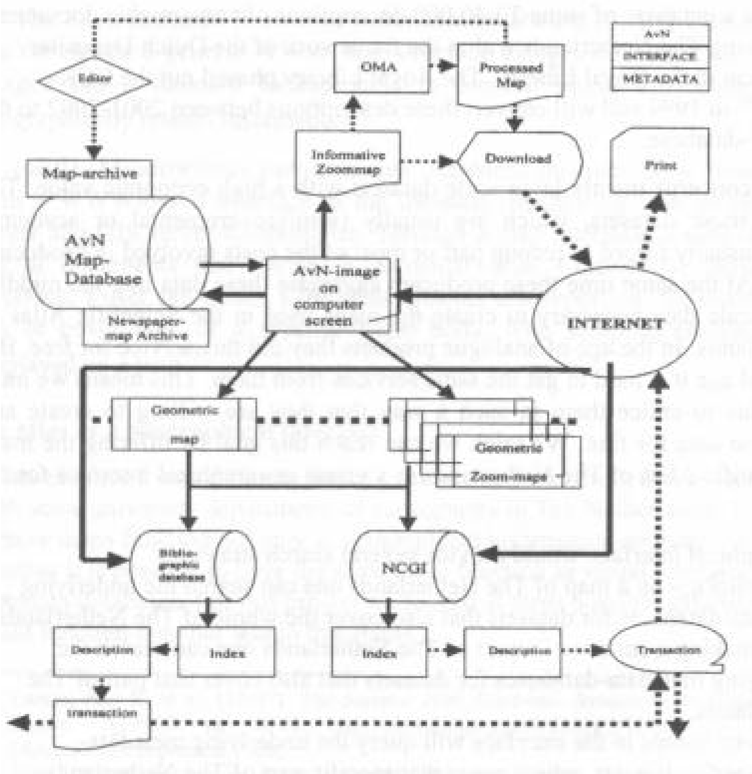


Diagram of a visual geographic interface

We try to sell the idea to the producers as follows. Is there a better way then the one described above to advertise the economically interesting datasets through such an interface and at the same time service education and the general cause! Because the Scientific Atlas of The Netherlands will not only function as a visual geographical interface, but at the same time will be an informational and

educational resource. The educational functions will be enhanced and become interactive when an Online Mapping Application (OMA) will be incorporated with which users can manipulate certain aspects of the maps offered. Maps produced with this application can be selected by an editor and added to an archive to serve as examples for future users.
At the same time some scanned samples of newspaper-maps have been added to the present database to get or give some insights in what way mass media use scientific mapping data to inform the public in general. In future digital maps from newspaper-archives could be downloaded to a sub database of the Scientific Atlas of The Netherlands and have the same function as maps added from the OMA-activities.

But why should we restrict ourselves to spatial metadata-databases for digital materials and not try to include metadata-databases for analogue materials. As long as the descriptions include geographical bounding-box co-ordinates the same kind of queries can be made on bibliographic databases as can be made on the NCGI-database. Because researchers require disparate sources to do their research, and because not all necessary information will be available on the Internet or in ready digital form, it would be a miss when bibliographical databases are left out of this scheme. The 30-40,000 descriptions of maps, which will be loaded into the PICA-database, contain all the necessary geometric or mathematical data necessary to use the query-functions as envisaged with the NCGI.
Furthermore the idea of one-stop-shop information gathering is so prevalent that we must do our utmost to realise this concept with the tools and all the data available.

Conclusion

In my opinion the metadata field will be expanding exponentially in contrast to the wish for basal descriptions as heralded in IFLA's *FRBR*[66]. But the burden of this growth will be put on the shoulders of the creators and the software programmers. Furthermore metadata will be used for a spectrum of functions, ranging from access to preservation. Once a metada-record has been created it will be constantly added to depending on the need for transformation to keep a DLO or database available for use. Cataloguers will become metadata information system managers and keep themselves busy with checking and validating the input. We also will create innovative user-interfaces which will replace the traditional listings. The most uncertain part of this set up is how the growth of databases will relate to the

[66] IFLA Study Group on the Functional Requirements for Bibliographic Records (1998). *Functional requirements for Bibliographic Records* (UBCIM publications - new series, Vol. 19). Also available through <http://www.ifla.org/VII/s13/frbr/frbr.pdf> [Accessed: 24 September 2000]

growth of DLOs, especially in the geospatial field. And in how far knowledge discovery and data mining, which are autonomous processes related to databases, will replace the need for metadata, except as internal checks and validations.

Jan Smits
Koninklijke Bibliotheek, National Library of The Netherlands
P.O. Box 90.407
NL-2509 LK Den Haag
The Netherlands
tel: +31 70 3140241
E-mail: jan.smits@kb.n