Strategies for Robust Digital Cartographic Steganography

Matthew T. Rice
National Center for Geographic Information and Analysis
Department of Geography
University of California, Santa Barbara
3611 Ellison Hall
Santa Barbara, CA 93106
USA
(805) 893-8652
(805) 893-8617 (fax)
rice@geog.ucsb.edu

Abstract

This research presents some new ideas for establishing and maintaining intellectual ownership rights for proprietary cartographic data through digital steganography, data alteration, and descriptive cataloging techniques. The ideas can be implement in both raster and vector data, and in some cases, the data identification techniques may be persistent under transformation, reprojection, resampling procedures, and in derivative or second-generation data.

Introduction

Digital cartographic databases present many new challenges for establishing and asserting intellectual property rights. Because digital data can be copied and disseminated easily, danger exists for misuse, misattribution, theft, and fraud. New methods for access and control are needed, particularly for governments, corporations, and individuals with proprietary, sensitive, classified, or expensive digital cartographic data. While the sharing of data and exchange of information is an ideal to strive for, the unauthorized and unscrupulous use of cartographic data inhibits those parties that would otherwise share data. Data creators deserve the right to maintain some control over their digital cartographic data, both while it is in their possession and after it has been released for use. Users of digital cartography data will also benefit if the origins, source, owner, and use restrictions for a piece of data are always clearly manifest. Respect for the intellectual property rights of digital cartographic data is a fundamental basis for the sharing of data and for establishing of the large shared digital geo-libraries of the future.

In previous eras, cartographic databases were kept as drawings on paper, mylar, or photographic film. The databases could be expropriated and copied, but only by tedious tracing and manual duplication. The paper, mylar, and film formats offered little possibility for data marking except for implanting simple cartographic traps. Most modern cartographic databases are kept in digital format, where data can be easily copied and disseminated over networks. The possibility for data expropriation is far greater, but the range of techniques for 'fingerprinting' and identifying this digital cartographic data are greater and more powerful.

Creating robust digital watermarks or fingerprints in cartographic data involves making slight alterations to the data's geometry, raster data values, coordinate floating point digits, object attributes, etc. . . and cataloging the presence and nature of such artifacts. The cartographic nature of the data makes this process more effective than in the standard implementation domain of photographic images, because these alterations or fingerprints can be implanted to take advantage of the geographic coordinates, geometry, and meaning of the data to make them more robust. Even the most clever digital fingerprints or watermarks may disappear or be lost through repetitive transformation and alteration, so a coordinated effort to create comprehensive descriptive catalogs of valuable data is suggested. Creating a comprehensive descriptive metadata catalog is more time and resource intensive than implanting a digital watermark, but comprehensive metadata provides a better means of identifying proprietary cartographic data after repetitive transformations and alterations have been made. Use of comprehensive descriptive metadata along with well-placed digital watermarks can provide a robust way of identifying and maintaining intellectual rights to digital cartographic data.

Many terms and words are used to describe the intentional alteration of data in order to hide a message or provide a means of authentication, including but no limited to: fingerprinting, data hiding,

data alteration, data profiling, watermarking, and steganography[1]. In a very general way, the goal of these techniques is to embed messages or mark data in an identifiable (but not necessarily visible) way and to provide a means of detecting the difference between one's own data and other datasets of the same area and theme. One of the most powerful of these techniques, digital steganography, consists of hiding a message inside a host document so that the communication of that message is not perceived. Substantial progress has been made in increasing the size and complexity of messages embedded using digital stenography. At present it is possible to imperceptibly hide video-in-video and image-in-image where the embedded message is 25% the size of the host (Chae and Manjunath, 1998). It is also reported to be possible to recover embedded images even when the original host document is not present (Chae and Manjunath, 1999). Digital watermarking is a similar technique but often has a primary purpose of authentification rather than hidden communication, and the watermarks are often visible rather than hidden, although they appear in both forms.

In the cartographic setting, digital steganography can be applied in the raster and vector realms, and when combined with ciphering techniques and comprehensive metadata cataloguing, digital stenography can be very robust in it's ability to communicate a hidden message for the purposes of establishing and asserting intellectual property rights.

This paper presents some settings for digital steganography in the cartographic realm, and illustrates its potential usefulness for both raster and vector data, particularly when combined with other techniques such as ciphering and metadata catalogs. Finally, a discussion of identifying derivative cartographic products will illustrate that although many techniques fail in embedded messages to derivative products, a coordinated, multi-facetd approach will offer the best chances at preserving a message to use in identifying one's own digital cartographic data.

Embedding Messages in Raster Cartographic Data

For raster data, the techniques of digital steganography are the similar to the non-cartographic realm, where they are seen most often in proprietary digital photograph archives. Although we are accustomed to thinking of watermarks as visible artifacts, digital watermarks can be both visible and invisible, with the invisible variety having wider application (figure 1). In the right-hand image, the invisible watermark embedded using a Digimarc® watermarking algorithm contains more information than the watermark in the left-hand image; namely, the identity of the creator, the year of copyright, and the public access level. Digimarc® watermarks can reportedly be traced through a watermarked image that is printed and scanned back into digital form.

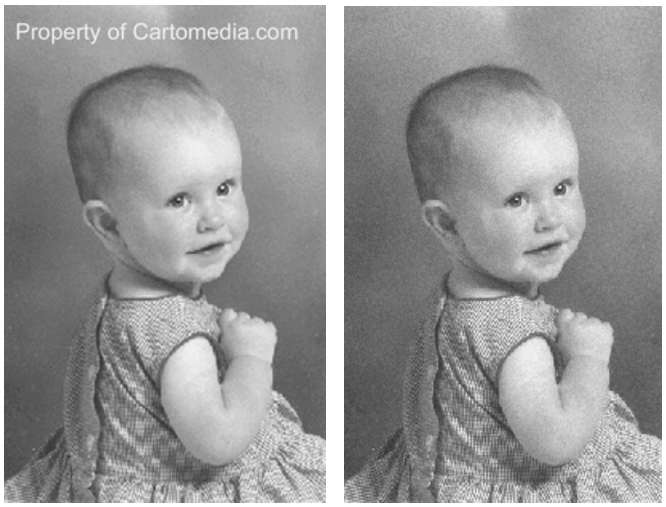

Figure 1. Picture with limited visible watermark (left) and picture with invisible digitial watermark containing creator identification, copyright year, and public access level (right).

Figures 2 and 3 show an aerial photograph from the Alexandria Digital Library at UC Santa Barbara (http://www.alexandria.ucsb.edu). This digital library contains hundreds of thousands of cartographic items as well as one of the worlds largest and most comprehensive digital gazetteers. These items are

[1] For more precise definitions and usages, see Katzenbeisser and Petitcolas, 2000 as well as the UCSB image processing and vision research lab web site, http://www-iplab.ece.ucsb.edu.

largely kept in electronic form, and are available to the public over the internet. Realizing that intellectual property rights and security would be an important issue in such a large digital library, time and effort was invested in determining whether or not state-of-the-art digital steganographic techniques could be employed to embed the Alexandria Digital Library logo (figure 2, signature image) into the digital aerial photographs.[2] As can be seen, the logo can be embedded and extracted, even under conditions such as lossy compression and resampling. The application of digital watermarking and steganographic techniques to raster cartographic data is no different. The same compression techniques used with these aerial photographs has been explored as a useful technique for compressing and storing digital elevation models (Shortridge 2000), and the techniques for extracting digital signature images would be the same for raster cartographic data. In addition to logos and trademarks, text can be embedded into raster structures. For digital cartographic data this could include identification of the author, creator, or owner; copyright information, use restrictions, change history, reference datums, spheroid, projection, etc. . . As with the embedded Alexandria Digital Library logo, this information could be extracted after lossy compression and other data-altering procedures.

Figure 2. Original Aerial Photo (left) and signature image (right)

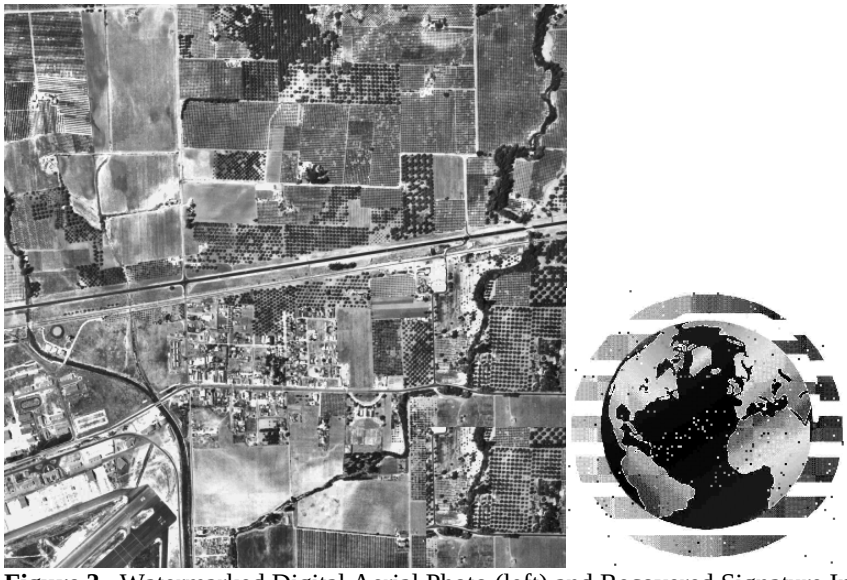

Figure 3. Watermarked Digital Aerial Photo (left) and Recovered Signature Image after 74% JPEG lossy compression

[2] See The UCSB Image Processing and Vision Research Lab web page, http://www-iplab.ece.ucsb.edu, for details of this application. Images used with permission, courtesy Dr. B.S. Manjunath, manj@ece.ucsb.edu.

For floating point grids and images, the embedded message text can be added as ASCII number sequences to any excess of non-significant cell or pixel values. For integer grids, the ASCII number sequences could be added to cell values if the range of values is large enough, or as an additive factor for groups of cells, as with the zonal grid functions implemented in ESRI's ArcInfo GIS software. The message text, if encrypted and turned into an ASCII number sequence, could discretely be added to the image header in a comment field or as an item in the attribute table. It should be stated that with each instance of embedded messages, a separate, secure metadatabase would hold the information about the nature and position of the embedded information.

Embedding Messages in Vector Cartographic Data

With vector data, adding information to data is easier, due to the richer information content. Hiding cipher text inside a piece of digital cartographic data is much easier when the data has associated derived geometric characteristics (length, width, distance, angle, aspect, etc. . .), coordinates, and associated attributes in a feature attribute table. The information contained in the coordinates can be analyzed to determine the relevancy and information content of the digits (Clarke and Battersby 2000), and information can be added to the low-information portion of the coordinates as ASCII number sequences or as cipher text turned into ASCII numbers. If an individual wanted to add identifying information to vector data, it would be as easy as translating this information to a number sequence, and adding it to non-significant digits of the coordinates for neighboring nodes along an arc. This identifying information could be retrieved in pieces or with portions missing, and be reassembled and translated using frequency analysis or other crypto-analytic methods. Rather than text, the embedded information could be an ID number referring to an entry in a separate metadata database. Derived geometric information from vector data (i.e., length, width, distance, angle, and direction) provides another place for embedded messages that is one level removed from the coordinate digits. The length of a line or angle between two lines could be altered to produce a specific number upon measurement, a portion of which translates into an ID or into a portion of a message.

Ciphering Techniques

Another data hiding technique with potential for increasing the security of digital steganography is cryptography. The standard alphabetic translation tool known as the Caesar shift, where messages are encoded using an cipher alphabet that is shifted one unit away from the standard alphabet (a → b, b → c, c → d, etc. . .), where "I love maps" turns into "J mpwf nbqt". These ciphers are relatively easy to break using visual inspection and guesses. When a single letter word appears, it is either an 'a' or an 'i'. Looking at the other letters of the cipher text and translating leads to a quick solution. The Caesar shift idea can be radically expanded using a tool developed in the sixteenth century by French diplomat and amateur cryptographer Blaise de Vigenere . Vigenere built on the works of several previous cryptographers and created a method for writing codes known as the Vigenere cipher. The cipher uses a square, which contains several cipher alphabets configured using Caesar shifts of increasing length (see figure 4). The first alphabet (row 1) is associated with the letter 'b' and translates 'a' into 'b', 'b' into 'c', 'c' into 'd', and so forth. The text of a message is encrypted using many simultaneous cipher alphabets, determined through the use of a keyword, such as "Mercator"[3]. The rows associated with this keyword are highlighted in the Vigenere square in figure 5. Using the "Mercator" keyword, a text message "Property of the US Geologic Survey" can be transformed into encrypted text as in figure 6. The cipher alphabet is changed after every letter of the Message text is encrypted. If the keyword for the Vigenere cipher is long and random, many of the typical crypto-analytic tools such as frequency analysis and are rendered useless, and this relatively simple technique becomes very powerful. In fact, Singh (1999, 122) suggests that the when the keyword for a Vigenere-type cipher is completely random and as long as the message text, it can be mathematically proven to be absolutely unbreakable. Several free programs exist on the web for implementing Vigenere ciphers and many of today's powerful ciphering and security systems are build upon the idea. Ciphering can provide an extremely high degree is security if implemented properly, and can be an excellent tool for implementing robust digital stenography for cartographic data.

[3] For an excellent and readable history of this technique and other cryptographic techniques, see Simon Singh, 1999. The Vigenere cipher is described in chapter 2.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
1	b	c	d	e	f	g	h	I	j	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z	a
2	c	d	e	f	g	h	I	j	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z	a	b
3	d	e	f	g	h	I	j	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z	a	b	c
4	e	f	g	h	I	j	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z	a	b	c	d
5	f	g	h	I	j	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z	a	b	c	d	e
6	g	h	I	j	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z	a	b	c	d	e	f
7	h	I	j	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z	a	b	c	d	e	f	g
8	I	j	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z	a	b	c	d	e	f	g	h
9	j	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z	a	b	c	d	e	f	g	h	I
10	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z	a	b	c	d	e	f	g	h	I	j
11	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z	a	b	c	d	e	f	g	h	I	j	k
12	m	n	o	p	q	r	s	t	u	v	w	x	y	z	a	b	c	d	e	f	g	h	I	j	k	l
13	n	o	p	q	r	s	t	u	v	w	x	y	z	a	b	c	d	e	f	g	h	I	j	k	l	m
14	o	p	q	o	s	t	u	v	x	x	y	z	a	b	c	d	e	f	g	h	I	j	k	l	m	n
15	p	q	r	s	t	u	v	w	x	y	z	a	b	c	d	e	f	g	h	I	j	k	l	m	n	o
16	q	r	s	t	u	v	w	x	y	z	a	b	c	d	e	f	g	h	I	j	k	l	m	n	o	p
17	r	s	t	u	v	w	x	y	z	a	b	c	d	e	f	g	h	I	j	k	l	m	n	o	p	q
18	s	t	u	v	w	x	y	z	a	b	c	d	e	f	g	h	I	j	k	l	m	n	o	p	q	r
19	t	u	v	w	x	y	z	a	b	c	d	e	f	g	h	I	j	k	l	m	n	o	p	q	r	s
20	u	v	w	x	y	z	a	b	c	d	e	f	g	h	I	j	k	l	m	n	o	p	q	r	s	t
21	v	w	x	y	z	a	b	c	d	e	f	g	h	I	j	k	l	m	n	o	p	q	r	s	t	u
22	w	x	y	z	a	b	c	d	e	f	g	h	I	j	k	l	m	n	o	p	q	r	s	t	u	v
23	x	y	z	a	b	c	d	e	f	g	h	I	j	k	l	m	n	o	p	q	r	s	t	u	v	w
24	y	z	a	b	c	d	e	f	g	h	I	j	k	l	m	n	o	p	q	r	s	t	u	v	w	x
25	z	a	b	c	d	e	f	g	h	I	j	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y
26	a	b	c	d	e	f	g	h	I	j	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z

Figure 4. Vigenere square for English alphabet

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
1	b	c	d	e	f	g	h	I	j	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z	a
2	**c**	d	e	f	g	h	I	j	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z	a	b
3	d	e	f	g	h	I	j	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z	a	b	c
4	**e**	f	g	h	I	j	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z	a	b	c	d
5	f	g	h	I	j	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z	a	b	c	d	e
6	g	h	I	j	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z	a	b	c	d	e	f
7	h	I	j	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z	a	b	c	d	e	f	g
8	I	j	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z	a	b	c	d	e	f	g	h
9	j	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z	a	b	c	d	e	f	g	h	I
10	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z	a	b	c	d	e	f	g	h	I	j
11	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z	a	b	c	d	e	f	g	h	I	j	k
12	**m**	n	o	p	q	r	s	t	u	v	w	x	y	z	a	b	c	d	e	f	g	h	I	j	k	l
13	n	o	p	q	r	s	t	u	v	w	x	y	z	a	b	c	d	e	f	g	h	I	j	k	l	m
14	**o**	p	q	o	s	t	u	v	x	x	y	z	a	b	c	d	e	f	g	h	I	j	k	l	m	n
15	p	q	r	s	t	u	v	w	x	y	z	a	b	c	d	e	f	g	h	I	j	k	l	m	n	o
16	q	r	s	t	u	v	w	x	y	z	a	b	c	d	e	f	g	h	I	j	k	l	m	n	o	p
17	**r**	s	t	u	v	w	x	y	z	a	b	c	d	e	f	g	h	I	j	k	l	m	n	o	p	q
18	s	t	u	v	w	x	y	z	a	b	c	d	e	f	g	h	I	j	k	l	m	n	o	p	q	r
19	**t**	u	v	w	x	y	z	a	b	c	d	e	f	g	h	I	j	k	l	m	n	o	p	q	r	s
20	u	v	w	x	y	z	a	b	c	d	e	f	g	h	I	j	k	l	m	n	o	p	q	r	s	t
21	v	w	x	y	z	a	b	c	d	e	f	g	h	I	j	k	l	m	n	o	p	q	r	s	t	u
22	w	x	y	z	a	b	c	d	e	f	g	h	I	j	k	l	m	n	o	p	q	r	s	t	u	v
23	x	y	z	a	b	c	d	e	f	g	h	I	j	k	l	m	n	o	p	q	r	s	t	u	v	W
24	y	z	a	b	c	d	e	f	g	h	I	j	k	l	m	n	o	p	q	r	s	t	u	v	w	X
25	z	a	b	c	d	e	f	g	h	I	j	k	l	m	n	o	p	q	r	s	t	u	v	w	x	Y
26	**a**	b	c	d	e	f	g	h	I	j	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y	Z

Figure 5. Vigenere square with cipher alphabets select for keyword ‘mercator’

Keyword	M	E	R	C	A	T	O	R	M	E	R	C	A	T	O	R	M	E	R	C	A
Message Text	p	r	o	p	e	r	t	y	o	f	t	h	e	u	s	g	e	o	l	o	g	i	c	s	u	r	v	e	y
Cipher Text	B	V	F	R	E	K	H	P	A	J	K	J	E	N	G	X	Q	S	C	Q	G	B	Q	J	G	V	M	F	Y
ASCII number	**66**	**86**	**70**	**82**	**69**	**75**	**72**	**80**	**65**	**74**	**75**	**74**	**69**	**78**	**71**	**88**	**81**	**83**	**67**	**81**	**71**	**66**	**81**	**74**	**71**	**86**	**77**	**70**	**86**

Figure 6. Keyword and message text with cipher text and ASCII number equivalents

Data Alteration, Descriptive Cataloging, and Detection in Derivative Products.

One of the easiest ways to identify a piece of digital cartographic data is to know it well. Humans are innately able to remember subtle differences in the human face, but when it comes to subtle differences between data sets, the skill is absent to a large degree. An individual may be able to easily pick their child out of a crowd based simply on the child's mannerisms and movement characteristics, but perhaps only a seasoned and veteran airphoto interpreter could detect individual differences in similar cartographic data upon cursory inspection.

For digital cartographic data, an easy way to provide a lasting characterization for identification purposes is to keep a comprehensive metadata database. If, for a digital elevation model, the user recorded the precise geographic locations of the higher order surface features such as pits and pinnacles, this information would be able to provide the key in identifying a piece of cartographic data after it is disseminated. Analytical information such as trend surface models for specific quadrants could be useful, as could statistical information such as distributions of cell values reflecting the eccentricities and individuality of the dataset. Even though messages embedded using steganography techniques could disappear during reprojection, resampling, or transformation, comprehensive statistical and analytical descriptions of the dataset could persist through derivative products. A hillshade derived from a digital elevation model would contain a level of patchiness corresponding to a statistical profile of a specific source digital elevation model stored in a metadata database. With this information, the source of a derivative cartographic product could be identified.

Another technique for 'marking' a digital cartographic dataset is to alter objects appearing in the dataset. At the Los Alamos National Laboratory (LANL), an extremely high-resolution digital elevation model is used to trace contaminant flows through the canyons in the area. The locations and heights of buildings can clearly be seen in the DEM and associated contours. An easy way to mark the dataset would be to adjust the heights of some non-essential buildings by two feet, effectively changing the way in which the dataset reflects the true landscape, but not in a way as to compromise the datasets utility in tracing contaminants. The presence of such altered artifacts could be cataloged in a metadata database, and from that point forward the valuable digital elevation model would be easily identified because of the alterations. In addition, changes to the dataset, such as those recorded in the Arcinfo GIS .log file can be encrypted and embedded using digital steganography. If LANL or other entity is afraid of sensitive or valuable data being leaked to the public by an employee, the change history and data user's name could be embedded within the digital elevation model as it is used. If the data is found in the hands of an unauthorized third party, the data could be analyzed to determine the last employee to access that data. Employee and general public knowledge about the use of such a technique could have a useful deterrent effect.

In the vector realm, several data alteration techniques have been suggested, such as inserting false streets and changing the orientation or length of an existing street. Even in GPS-enabled vehicles, altering the position of a street by two meters in a digital cartographic database, or changing the amount of curvature slightly would not be noticeable and would not affect the usability and utility of the dataset. The locations of intersections on street networks from different data vendors are reported to differ by as much as 200 meters (Noronha and Goodchild, 2000). A two meters shift in the location of a road or a slight change in the curvature would make the data easily identifiable and would not detract from its utility as a reference for vehicle navigation systems in any way.

In this vein, Rice (1998) analyzed the positional discrepancies between three different topographic bases for the same area in an effort to create a visualization based conflation algorithm to correct for relative positional error. The three different topographic bases differed in the position of key features by as much as 50 meters. Shifting the location of topographic features by small amounts would provide a measure of uniqueness for identification purposes and in no way keep the dataset from being useful as a reference.

Conclusion

Digital steganography can provide a robust means of embedding identification information into both raster and vector digital cartographic datasets. This information can be extracted at a later time and used to identify the datasets owner, copyright status, access and use restriction, cost, etc. . . The technique of digital steganography is traditionally applied in the raster domain, but it can be powerful in the vector domain, using coordinate digits as a host. In this context, information can be

stored directly in the coordinate digits or can appear through measurements derived from coordinates such as length, width, direction, and distance. Clarke and Battersby (2000) have provided a starting point for this technique through their analysis of the information content of coordinate digits. When digital steganography is combined with several other methods of 'marking' a dataset such as alteration, the dataset becomes even more identifiable. The use of metadata catalogs to store the unique and identifying characteristics of a dataset provides a third and important method for identification of a digital cartographic resource. The coordinated application of these approaches can lead to digital cartographic data that is personalized and clearly identifiable as a person's intellectual property.

References

Chae, J.J., and B. S. Manjunath, "A Robust Embedded Data from Wavelet Coefficients", Proc.SPIE: Storage and Retrieval for Image and Video Databases VI, vol. 3312, pp. 308-317, San Jose, CA, Jan 1998.

Chae, J.J., and B.S. Manjunath. "A technique for image data hiding and reconstruction without host image", Proceedings of the SPIE - The International Society for Optical Engineering, vol.3657,(Security and Watermarking of Multimedia Contents), San Jose, CA, USA, pp.386-96, Jan 1999.

Clarke, Keith C, and Sarah E. Battersby. "The Coordinate Digit Density function for map information content analysis". In Preliminary Program, 2001 ACSM-CLSA-NALS-WFPS Conference and Exposition, Las Vegas, NV, March 19-21. Gaithersburg, Maryland: American Congress on Surveying and Mapping, (2001).

Katzenbeisser, Stefan, and Fabien A. P. Petitcolas, eds. Information Hiding Techniques for Steganography and Digital Watermarking. Boston, Massachusetts: Artech House, 2000.

Noronha, Val, and Goodchild, Michael F. "Map Accuracy and Location Expression in Transportation – Reality and Prospects." Transportation Research Part C 8 (2000), 53-69.

Shortridge, Ashton. "Compact Data Models for Spatially Continuous Phenomena", In GIScience 2000, The First International Conference on Geographic Information Science, Savannah, Georgia, October 28 – 31, 2000.

Singh, Simon. The Code Book: The Evolution of Secrecy from Mary Queen of Scots to Quantum Cryptography. New York: Random House, 1999.