The Trustworthiness of Digital Records

Luciana Duranti
Professor, Archival Studies
University of British Columbia

Education Code: FR03-2184

Learning Objectives

Upon completion of this session, participants will be able to:

1. Identify records in digital systems
2. Determine when a digital system should contain records, but does not
3. Propose ways of modifying a system to ensure it does create records
4. Develop functional requirements and procedures that ensure digital records are reliably created and maintained accurately and authentically
5. Assess whether a recordkeeping system can be regarded as having integrity

Most Important Issues

- **Conceptual**
  - The Concept of Record
  - The Concept of Trustworthiness
  - The Concept of Life Cycle
- **Methodological**
  - Appraisal
  - Preservation
- **Strategic**
  - Relationship Creator-Preserver
  - The Role of the Record Professional

The Concept of Record

- **Record**: any document made or received by a physical or juridical person in the course of activity as an instrument and by-product of it, and kept for action or reference
- **Document**: recorded information (i.e., information affixed to a medium in an objectified and syntactic form)
- **Information**: "intelligence given," or a message intended for communication across time and space
- **Data**: the smallest meaningful piece of information

Digital Record Characteristics

- **Act**: an action in which the records participates or which the record supports (naturalness and impartiality)
- **Persons Concurring to Its Creation**: author, writer, originator, addressee, and creator
- **Archival Bond**: explicit linkages to other records inside or outside the system (uniqueness)
- **Identifiable Contexts**: juridical-administrative, provenancial, procedural, documentary, technological (interrelatedness)
- **Medium**: necessary part of the technological context, not of the record
- **Fixed Form and Stable Content**

Fixed Form

* An entity has fixed form if its binary content is stored so that the message it conveys can be rendered with the same documentary presentation it had on the screen when first saved (different digital presentation: Word to .pdf)
* An entity has fixed form also if the same content can be presented on the screen in several different ways in a limited series of possibilities: we have a different documentary presentation of the same stored record having stable content and fixed form (e.g. statistical data viewed as a pie chart, a bar chart, or a table)

Stable Content

* An entity has stable content if the data and the message it conveys are unchanged and unchangeable, meaning that data cannot be overwritten, altered, deleted or added to

* **Bounded Variability**: when changes to the documentary presentation of a determined stable content are limited and controlled by fixed rules, so that the same query or interaction always generates the same result, and we have different views of different subsets of content, due to the intention of the author or to different operating systems or applications

Digital Record Characteristics (cont.)

- **Formal Elements**: constituent parts of the record documentary form as shown on its face, e.g. address, salutation, preamble, complimentary close
- **Metadata**: the attributes of the records that demonstrate its identity and integrity
- **Digital Components**: stored digital entities that either contain one or more records or are contained in the record and require a specific preservation measure

Stored and Manifested Records

- **Stored record**: it is constituted of the digital component(s) used in re-producing it, which comprise the data to be processed in order to manifest the record (content data and form data) and the rules for processing the data, including those enabling variations (composition data)

- **Manifested record**: the visualization or instantiation of the record in a form suitable for presentation to a person or a system. Sometimes, it does not have a corresponding stored record, but it is re-created from fixed content data when a user's action associates them with specific form data and composition data (e.g. a record produced from a relational database)

Types of Digital Records

Static: They do not provide possibilities for changing their manifest content or form beyond opening, closing and navigating: e-mail, reports, sound recordings, motion video, snapshots of web pages

Interactive: They present variable content, form, or both, and the rules governing the content and form of presentation may be either fixed or variable

Interactive Entities

- **Non-dynamic**: the rules governing the presentation of content and form do not vary, and the content presented each time is selected from a fixed store of data. Ex. Interactive web pages, online catalogs, records enabling performances—they are records
- **Dynamic**: the rules governing the presentation of content and form may vary—they are either information systems or potential records

Interactive Information Systems

* Entities produced in dynamic computing applications that select different sets of rules to produce documents, depending on user input, sources of content data, and characteristic of content (weather sites)
* Entities produced by evolutionary computing where the software generating them can change autonomously (scheduling and modeling of financial markets; edutainment sites)

Interactive Potential Records

- Entities where the variation is due to data that change frequently, because the design permits updating, replacement or alterations; allows data collection from users or about user interactions or actions; or uses these data to determine subsequent presentations (e.g. Land Registry)
- Entities where the variation is due to data received from external sources and not stored within the system (e.g. GIS)

They are presently not records but should be made into records if they fulfill one of the records functions.

Records Functions
(the way a record relates to an action)

- ***Ad substantiam*** (dispositive, e.g., contracts)
- ***Ad probationem*** (probative, e.g., registries)
- **Supporting**: generated to be used in the course of activity (ies) as a source of information, often by multiple users (e.g., GIS)
- **Narrative**: generated on a purely discretionary basis only as a means of communication (e.g., most e-mails, memos, web sites)

Records Functions

- **Instructive**: provide guidance on the way in which external data or documents are to be presented (e.g., scores, scripts, regulations, manuals of procedure, instructions for filling out forms)
- **Enabling**: enable the performance of artworks (software patches), the execution of business transactions (interacting business applications), the conduct of experiments (a workflow generated and used to carry out an experiment of which it is instrument, byproduct and residue), the analysis of observational data (interpreting software), etc. Most of them are stored only records.

Digital Forensics View

- **Computer Stored**: They contain human statements and are considered hearsay (tested for truthfulness and accuracy under the business records exception to the hearsay rule): e.g. e-mail messages, word processing documents, and Internet chat room messages.
- **Computer Generated**: They do not contain human statements, but they are the output of a computer program designed to process input following a defined algorithm (tested for authenticity on the basis of the functioning of the computer program): e.g. server log-in records from Internet service providers, ATM records.
- **Computer Stored & Generated**: e.g. a spreadsheet record that has received human input followed by computer processing (the mathematical operations of the spreadsheet program).

Trustworthiness

Reliability

The trustworthiness of a record as a statement of fact,

based on:

- the competence of its author
- the controls on its creation

Accuracy

The correctness and precision of a record's content

based on:

- the competence of its author
- the controls on content recording and transmission

Authenticity

The trustworthiness of a record that is what it purports to be, untampered with and uncorrupted

based on:

- identity
- Integrity
- reliability of the system

Authenticity: Identity

The whole of the attributes of a record that characterize it as unique, and that distinguish it from other records.

Identity metadata:

- names of the persons concurring in its creation
- date(s) and time(s) of issuing, creation and transmission
- the matter or action in which it participates
- the expression of its archival bond
- documentary form
- digital presentation
- the indication of any attachment(s)
- digital signature
- name of the person responsible for the business matter

Authenticity: Integrity

A record has integrity if the message it is meant to communicate in order to achieve its purpose is unaltered.

Integrity metadata:

- name(s) of handling persons over time
- name of person responsible for keeping the record
- indication of annotations
- indication of technical changes
- indication of presence or removal of digital signature
- time of planned removal from the system
- time of transfer to a custodian
- time of planned deletion
- existence and location of duplicates outside the system

Authentication

A means of declaring the authenticity of a record at one particular moment in time -- possibly without regard to other evidence of identity and integrity.

Example: the digital signature. Functionally equivalent to medieval seals (not signatures):

* verifies origin (identity)
* certifies intactness (integrity)
* makes record indisputable and incontestable (non-repudiation)

The analogy is not perfect, because the medieval seal was associated exclusively with a person, while the digital signature is associated with a given person and a specific record, and because the former is an expression of authority, while the latter is only a mathematical expression

Trusted Systems

Rules, and tools and methods to implement rules, for

Making reliable and accurate records

- record-identity metadata schemes
- business and documentary procedures integrated in a workflow structure linked to classification schemes and filing plans
- specifications of record forms
- record-making access privileges

Maintaining and keeping authentic records

- record-integrity metadata schemes
- classification schemes and filing plans
- linked retention schedule
- registration system
- retrieval system
- record-keeping access privileges

Digital Forensics View

The Daubert rules are used for scientific and technical evidence (not substantive but demonstrative evidence—intention vs. capability):

- the theory, procedure or process for making or keeping the record has been tested or cannot be tampered with
- it has been subjected to peer review or publication
- the known or potential error rate is acceptable
- it is generally accepted within the relevant scientific community

The judiciary looks for **repeatability, verifiability, objectivity** and **transparency**

Case Study #1: the Alsace-Moselle Land Registry

The registry is required by the French real estate law, as the means to fulfill the requirement that the legal status of property (including the various forms of mortgages on the property) must be made publicly available to interested third parties by means of inscription within a land registry.

The Procedure of Inscription

1. An electronic request for inscription is generated by the notary using custom software, which connects to the land registry in order to retrieve the information relative to the parties or parcels
2. Once the request is received at the land registry office, it is dated. This date determines the inception of the rights on the property.
3. For each request, an electronic file is created containing all of the associated documents (contract, cadastre, etc.), as scanned imaged files where they do not exist as digital data sets to which the request can be linked

The Procedure of Inscription (cont.)

4. A **draft order of inscription** is prepared. Inscriptions are also drafted directly in the database, but are not visible to outside users of the database until a judge has signed them; the draft order is transferred to the judge's "in box" in the form of an XML document
5. The judge is responsible for the required verifications; however, the custom software of the land registry office provides him/her with a **"before" and "after" view of the inscription**, that is, of the changes to the registry which the inscription will effect in the database;
6. After identifying himself through biometric (fingerprint) scan and inserting a smartcard with his private signature key, **the judge signs the draft order.** At that moment, **in a single step, the order is generated and signed, producing an inscription, and the relevant fields of the database are updated.**

Digital Entities in the Registry

- **The order**, which becomes the inscription, listing the information relative to the land parcel, the parties to the transaction, and the nature of the transaction. It is delineated in fields, using XML tags, and may thus be readily processed. It is authored by the judge, who dates and signs it.
- **The tables of a relational database** (i.e., one table that records the characteristics of land owners, another of land parcels, another of the *charges,* another of the mortgages); with **links between the tables** (using primary/foreign key mechanisms) that establish relationships between relevant data in the tables. The two most important views offered by the computerized land registry are (a) the ownership history of a given land parcel and (b) the set of land parcels owned by a particular individual.

Technological Structure

- An Oracle database, containing the land registry data;
- Personal (Windows) computers, for registry clerks, running web-based applications for consulting the registry and managing the inscription process;
- Plugs-ins for commercial notarial software for integration with the land registry;
- Personal computers, for land registry judges, running web-based applications for consulting the registry and for finalizing inscriptions to the registry and equipped with biometric identification peripherals, and digital signature software;
- A PKI infrastructure, linking together all land registry offices and the central database, so that judges may sign orders and add inscriptions to the registry.

What Is New?

- The system uses the digital signatures to provide continuous authentication services, that is, regularly performed declarations of the integrity and origin of the data.
- Digital signatures provide an extreme assessment of the integrity of data: if even a single bit of the signed data is modified, the signature fails.
- They also compare the orders with the inscriptions every time their authenticity is questioned.

Issues

- The District Archives must receive registration records after 5 years
- While the acquisition of the orders by the District Archives, as stand-alone documents, poses no particular problems, that of the inscriptions does
- The digital inscriptions are not records, the land registry as a whole is.
- As a record, the land registry cannot be understood outside of its dynamic and interactive capabilities.
- The inscriptions cannot be authenticated outside the PKI infrastructure
- Migration to overcome obsolescence risks loss of interoperability

Possible Solution

The definition of an XML schema which may serve as a translation device between the complex data model used by the land registry, and a less complex model, to be defined, sufficient to satisfy the needs of future users.

Inscriptions could then be exported to a file according to the XML schema and imported into relational database sufficiently simple to be maintained by the designated preserver (e.g., Microsoft Access).

Case Study #2: the VanMap

The cross-corporate GIS created by the City of Vancouver and used by staff in

- Engineering
- Planning
- Permits and Licenses
- By-law Enforcement
- Social Planning
- Police
- Fire and Rescue
- Parks and Recreation

http://vanmappub.vancouver.ca - VanMap Public ActiveX Viewer version for Windows - Microsoft Internet Explorer provided by ICBC

VanMap
public edition

Contact Us

Legend Options

Address Search

Options | Reset | Help

Number Street

or select E 1st Av Go

Toolbox

Right-Click Menu | Help

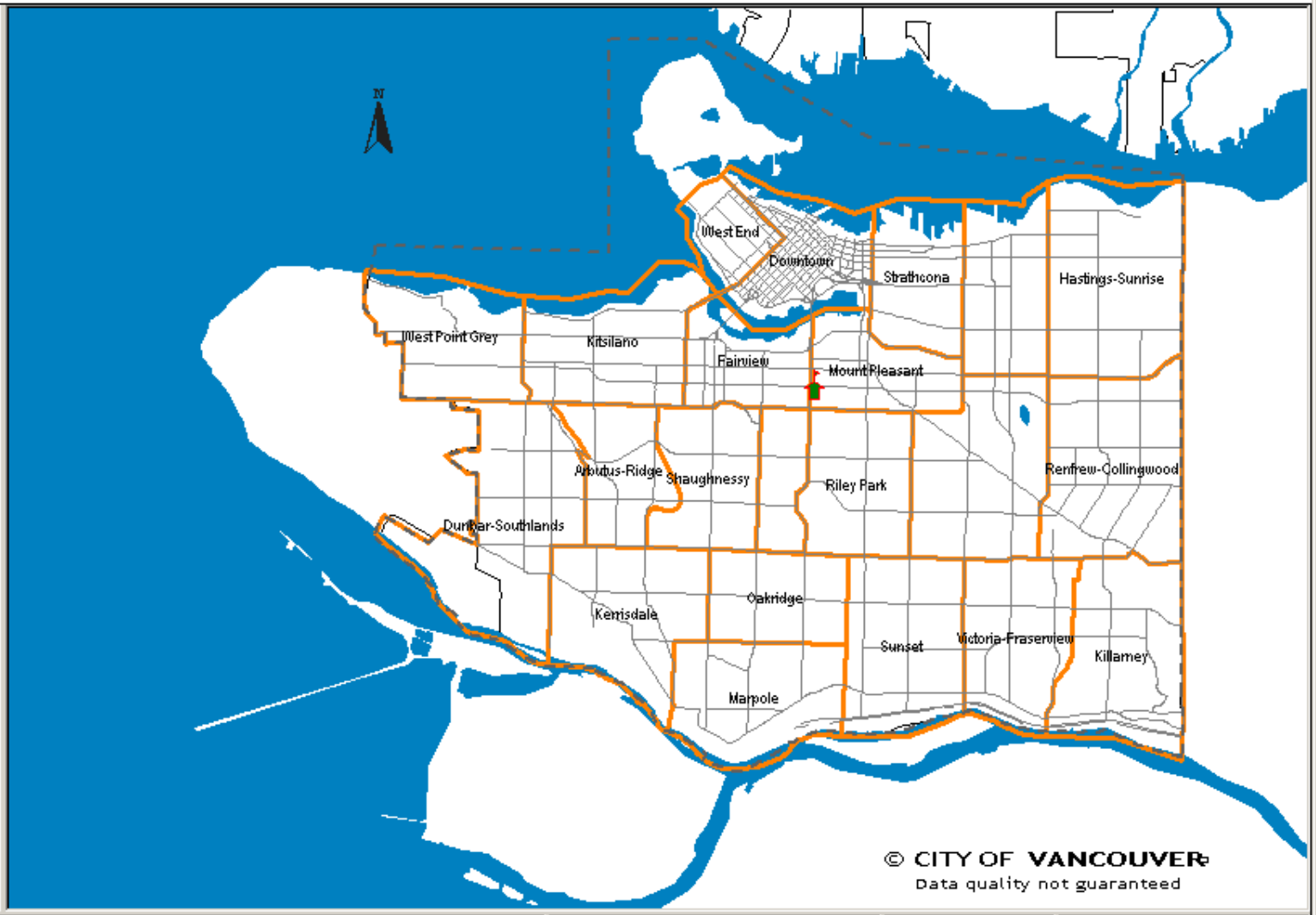
Help

About the Data

Applications

Select Application:

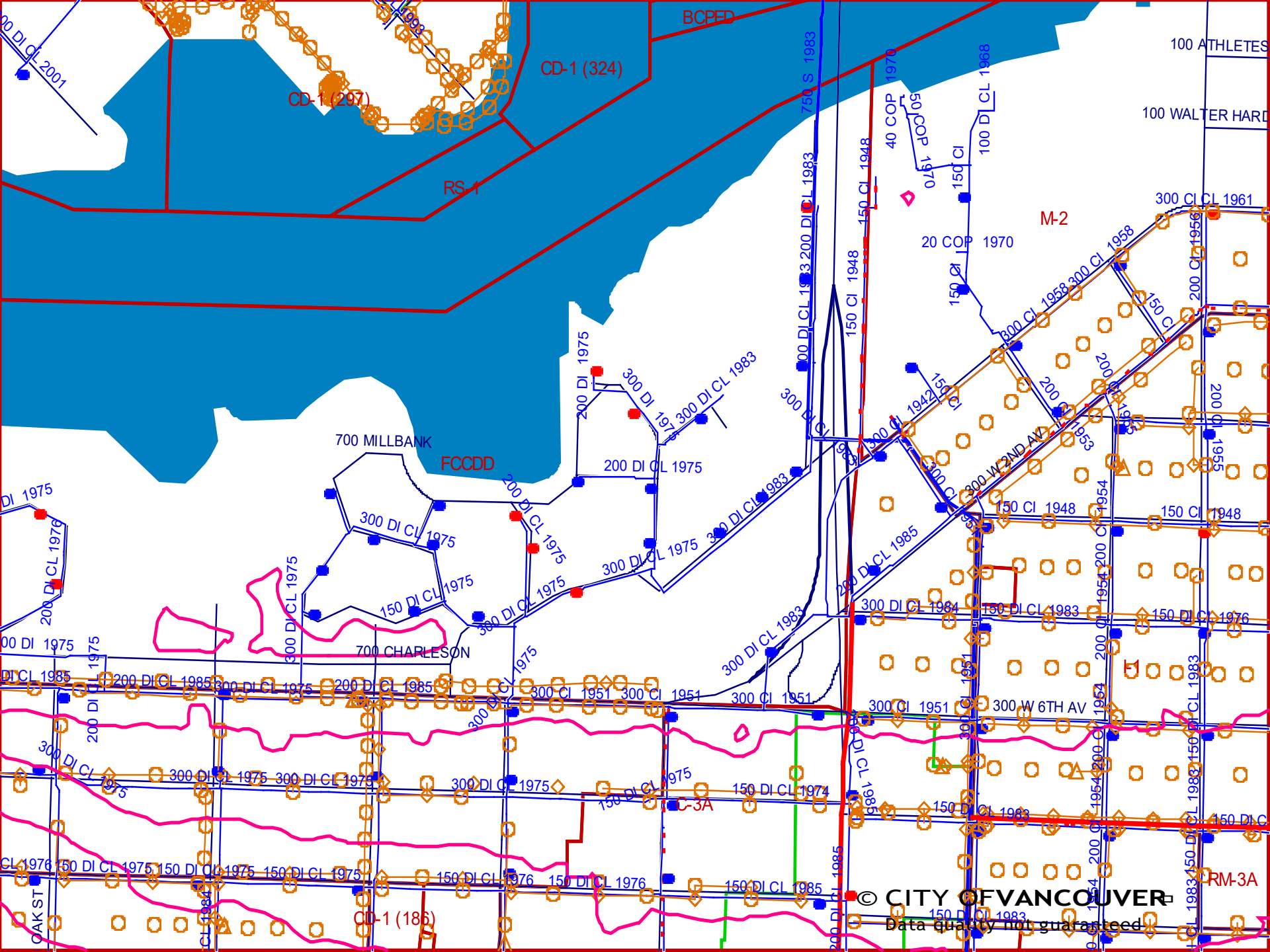
- Olympic Venues
- City Boundary
- Public Art
- Public Places
- The Road Ahead
- City Projects
- Child Care Facilities
- Web Cameras
- Non-Market Housing
- City-owned Properties
- Cemetery
- Urban Agriculture
- Contour Lines
- Truck Routes
- Canada Line Alignment
- Traffic Related
- Sewer
- Water
- City Streets Network
- View Cones
- Subdivision Categories
- DCL Areas
- Zoning Districts Types
- Zoning Districts
- Business Improvement Areas
- False Creek Navigable Chan
- Youth
- Property Information
- Facet Grids
- Administrative & Service Area
- Shore Lines (2002)
- Foreshore Contour Line
- Water Bodies

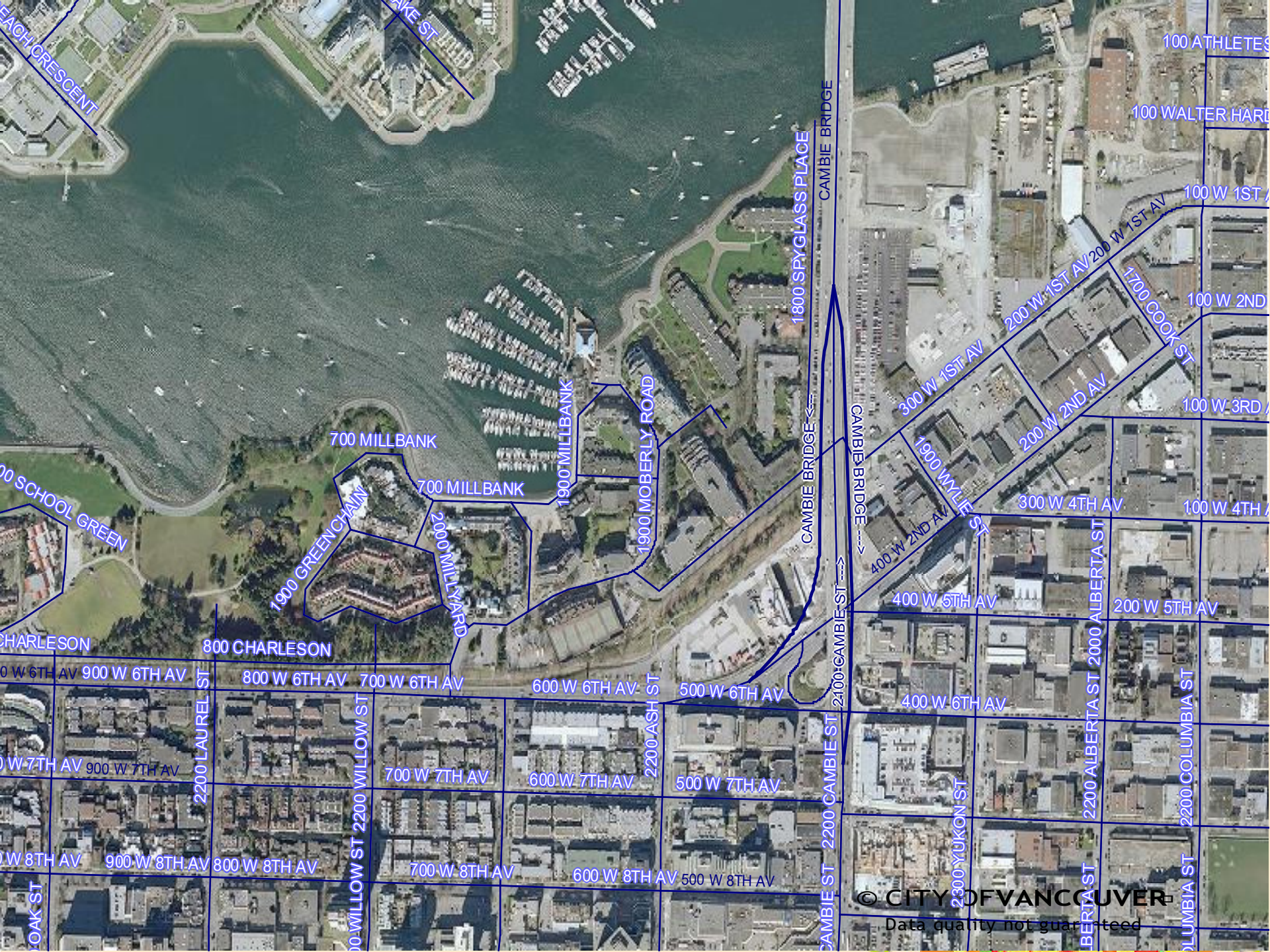


Water Bodies : Strait of Georgia | 0 feature(s) selected | 1 : 109,043 | 14.4 x 10.0 (mi)

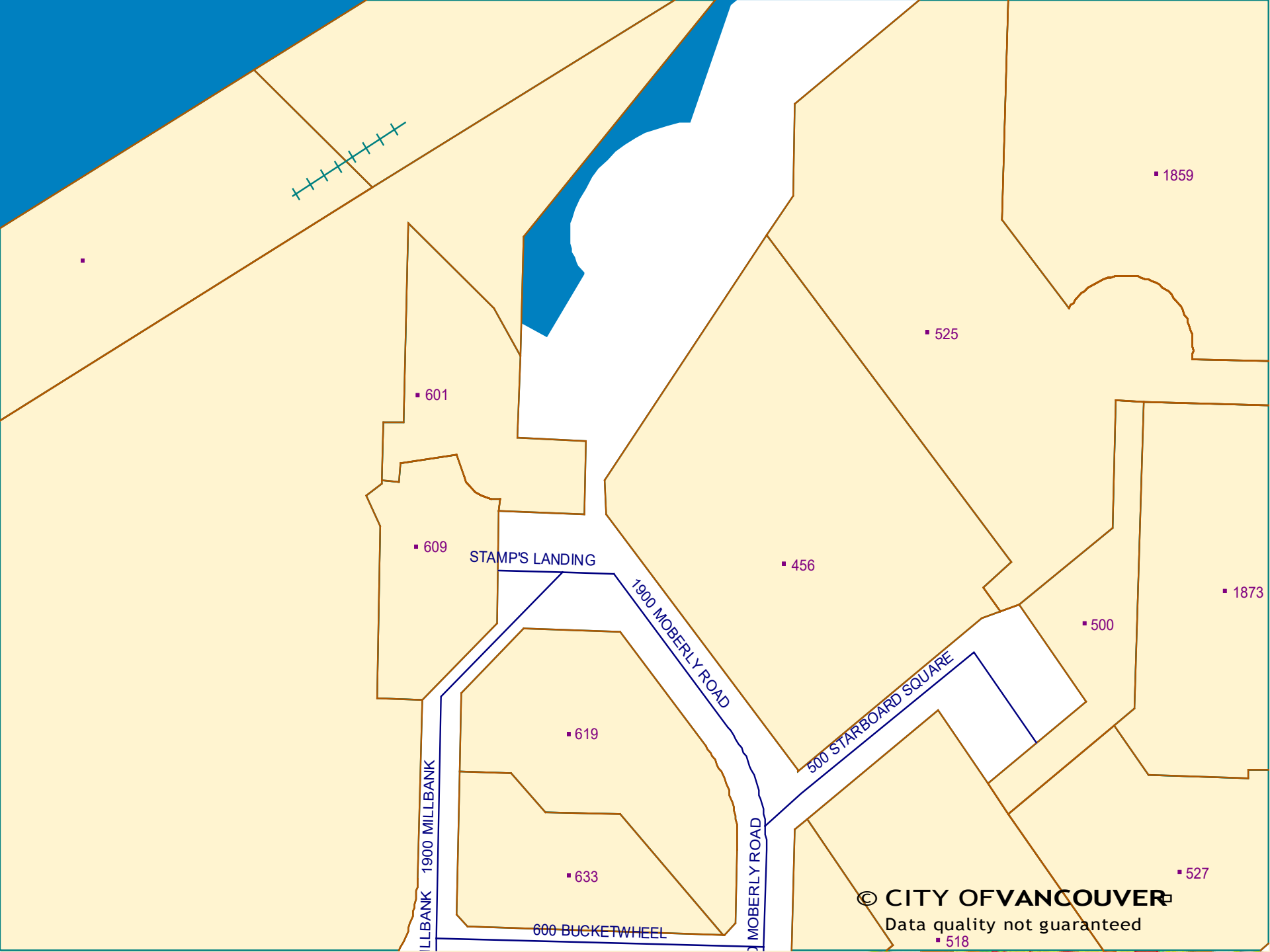
Internet

Start | Inbox - Micro... | 2 Windows ... | Microsoft Excel | Symposium 2... | Presentation1 | ACA presenta... | 2 Internet ... | 12:42 PM


BCPED
CD-1 (324)
CD-1 (297)
RS-1
M-2
C-3A
RM-3A
CD-1 (186)
FCCDD
700 MILLBANK
700 CHARLESON
100 ATHLETES
100 WALTER HARD
300 W 2ND AV
300 W 6TH AV
OAK ST
© CITY OF VANCOUVER
Data quality not guaranteed


EACH CRESCENT
AKE ST
1800 SPYGLASS PLACE
CAMBIE BRIDGE
100 ATHLETE S
100 WALTER HARD
100 W 1ST A
200 W 1ST AV
1700 COOK ST
100 W 2ND
300 W 1ST AV
200 W 2ND AV
100 W 3RD A
700 MILLBANK
700 MILLBANK
1900 MILLBANK
1900 MOBERLY ROAD
CAMBIE BRIDGE <---
CAMBIE BRIDGE <---
CAMBIE BRIDGE --->
1900 WYLIE ST
300 W 4TH AV
100 W 4TH A
400 W 2ND AV
00 SCHOOL GREEN
1900 GREENCHAIN
2000 MILLYARD
400 W 5TH AV
200 W 5TH AV
CHARLESON
800 CHARLESON
0 W 6TH AV
900 W 6TH AV
800 W 6TH AV
700 W 6TH AV
600 W 6TH AV
500 W 6TH AV
2400 CAMBIE ST
400 W 6TH AV
2200 ALBERTA ST
2000 ALBERTA ST
2200 COLUMBIA ST
0 W 7TH AV
900 W 7TH AV
2200 LAUREL ST
2200 WILLOW ST
700 W 7TH AV
600 W 7TH AV
2200 ASH ST
500 W 7TH AV
2200 CAMBIE ST
2300 YUKON ST
0 W 8TH AV
900 W 8TH AV
800 W 8TH AV
700 W 8TH AV
600 W 8TH AV
500 W 8TH AV
OAK ST
2200 WILLOW ST
CAMBIE ST
ALBERTA ST
COLUMBIA ST
© CITY OF VANCOUVER
Data quality not guaranteed


1859
525
601
609
456
1873
500
619
633
527
518
© CITY OF VANCOUVER
Data quality not guaranteed


1859
525
601
609
STAMP'S LANDING
456
1900 MOBERLY ROAD
1873
500
500 STARBOARD SQUARE
619
1900 MILLBANK
633
600 BUCKETWHEEL
MOBERLY ROAD
527
518
© CITY OF VANCOUVER
Data quality not guaranteed

VanMap
public edition

Print date: 13-Feb-07,01:52 PM

Tax Attributes Report

urce: The city's Property Tax System. (see About Data for the latest update.)

ecord	PID or Strata Plan Number	Assessment Roll Number (Folio)	From Address Number	To Address Number	Street Name	Property Postcode	Lot	Block	District Lot	Plan	Legal Description	Previous Year Land Value($)	Previous Year Improvement Value($)	Current Year Land Value ($)	Current Year Improvement Value($)	Asse
	VAS1831	164632770001		601	STAMPS LANDING	V5Z 3Z1	1		FC	VAS1831	LOT 1 PLAN VAS1831 DISTRICT LOT F C NEW WESTMINSTER AN UNDIVIDED 795 /808 SHARE IN THE COMMON PROPERTY T HEREIN EX PLAN 18514, 18515, & 1851 6.	570000	1124000	584000	1135000	2007
	VAS1831	164632770002		601	STAMPS LANDING	V5Z 3Z1	2		FC	VAS1831	LOT 2 PLAN VAS1831 DISTRICT LOT F C NEW WESTMINSTER AN UNDIVIDED 13/ 808 SHARE IN THE COMMON PROPERTY TH EREIN EX PLAN 18514, 18515, & 18516 .	1200	39600	7800	40000	2007

VanMap Technical Components

* Oracle Spatial database
* Other databases linked to it, existing in a variety of offices, of a variety of local authorities, and whose data flow continuously in the Oracle database
* CAD drawings, satellite imagery, photographs, html pages
* Autodesk MapGuide
* Autodesk ActiveX Viewer
* Application servers
* Web server

VanMap is a Dynamic Information System

- Data that often do not exist anywhere else, especially in the correlated form showed on the GIS layers, are overwritten without being saved
- The data are viewed as maps but these views are not saved
- New layers are being added all the time
- VanMap does not contain records

Is VanMap as a Whole a Record?

Yes!

* It is made and received in the course of a practical activity
* It is an instrument and a by-product of that activity
* It is an indivisible unit affixed to a medium
* It has all the diplomatic characteristics of a record in terms of persons and contexts

Is VanMap a Record?

No!

* It lacks fixed form and stable content
* It is not set aside for action or reference

Thus, it cannot be used to render an account of the decisions made or as a memorial.

Can VanMap Become a Record?

* Yes, if we introduce fixed form and stable content
* We need to configure the system so that, as each layer is updated, the data are saved rather than overwritten
* Then we need to develop a means of reproducing VanMap as it was on any given date

What About Taking Map Views?

- The preserver cannot do so or would become the creator of digital objects never used by the creator in the course of business, a creator of his/her own records
- It is not feasible to require City staff to save the map views in connection with the decisions based on them
- We have to preserve not what the staff member ***saw*** at a given point in time but what s/he ***would have been able to see***
- A detailed documentation of the business process would support this preservation activity

How to Build a GIS Preservation Environment

* Step 1: save the empty layers
* Step 2: add metadata to the layers
* Step 3: store the data in a secure environment
* Step 4: create infrastructure independence
* Step 5: migrate to new/neutral technology platforms
* Step 6: reproduce the system

Using Data Grid Technology

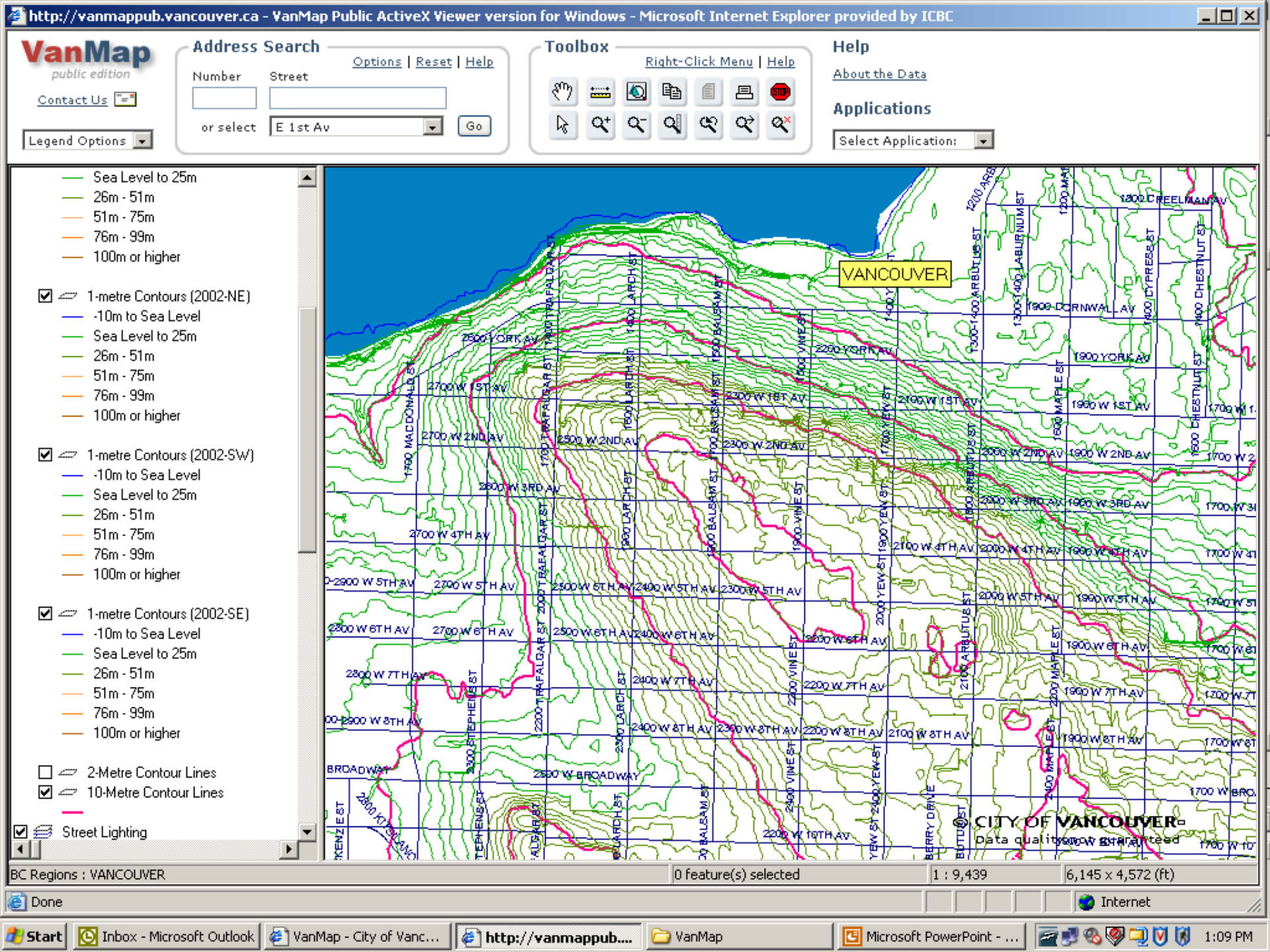
- Manages data and their associated metadata
- Separates the data from dependence on original creating infrastructure
- Maintains audit trails of all operations performed on the data
- Manages access and retrieval
- Supports migration of data to new platforms

Data Grid and VanMap

* Data grid is inserted between the data storage systems and the access applications
* Each saved layer within the GIS is independently registered in the data grid
* Queries based on the chronological date of the data are used to reproduce VanMap layers
- Queried data are loaded into a different GIS product

What Is Preserved?

* The data themselves
* The ability to see the data available on a given day and time
* The ability to render the data as interactive maps
* Presentation elements such as colours and fonts do not necessarily have to be preserved given the costs of doing so


http://vanmappub.vancouver.ca - VanMap Public ActiveX Viewer version for Windows - Microsoft Internet Explorer provided by ICBC
VanMap
public edition
Contact Us
Legend Options
Address Search
Options | Reset | Help
Number
Street
or select
E 1st Av
Go
Toolbox
Right-Click Menu | Help
Help
About the Data
Applications
Select Application:
Sea Level to 25m
26m - 51m
51m - 75m
76m - 99m
100m or higher
1-metre Contours (2002-NE)
-10m to Sea Level
Sea Level to 25m
26m - 51m
51m - 75m
76m - 99m
100m or higher
1-metre Contours (2002-SW)
-10m to Sea Level
Sea Level to 25m
26m - 51m
51m - 75m
76m - 99m
100m or higher
1-metre Contours (2002-SE)
-10m to Sea Level
Sea Level to 25m
26m - 51m
51m - 75m
76m - 99m
100m or higher
2-Metre Contour Lines
10-Metre Contour Lines
Street Lighting
VANCOUVER
© CITY OF VANCOUVER
Data quality not guaranteed
BC Regions : VANCOUVER
0 feature(s) selected
1 : 9,439
6,145 x 4,572 (ft)
Done
Internet
Start
Inbox - Microsoft Outlook
VanMap - City of Vanc...
http://vanmappub....
VanMap
Microsoft PowerPoint - ...
1:09 PM

The Same Solution for All GIS?

No!

- What is identified as **the record to be generated and maintained over time depends on the use of the data by the creator and the reason for having records** rather that fluid information.
- The research GIS of the **Archaeological Society of Arizona** requires preservation of its ability to make the users detect underground materials from the layers showing vegetation and stratifications of the soil and therefore preservation of the records that suggest that excavations should be carried out.
- The **Canadian Atlas of Antarctica** requires preservation of the content of external users interactions with specific layers and therefore preservation of the users' records that have changed the system output.

Why is It Important to Know What a Digital Record Is?

- Evidence
- Accountability
- Protection of Our Rights
- Preserving Our Identity
- Understanding the Past
- Relying on Our Sources
- ***Quod non est in actis, non est in mundo*** - What is not in the records does not exist

InterPARES Web Site

www.interpares.org

Digital Records Forensics Web Site

www.digitalrecordsforensics.org