A Case Study in e-Government: VanMap

Evelyn McLellan

Insurance Corporation of British Columbia

Glenn Dingwall

City of Vancouver Records & Archives

Reagan Moore

San Diego Supercomputer Center

Richard Marciano

Sustainable Archives & Library Technologies lab (SALT)

San Diego Supercomputer Center

UCSD

SDSC

CS24: Preservation of the City of Vancouver Geographic Information System (GIS): VanMap

- *Phase I: 2004-2005*
  - *Insurance Corporation of British Columbia:*
    - **Evelyn McLellan** (case study leader)
  - *UBC:*
    - **Luciana Duranti**, Director InterPARES2
    - **Eleanor Kleiber**, Research Assistant
    - **Catherine Miller**, Research Assistant
  - *City of Vancouver Records & Archives:*
    - **Glenn Dingwall**, Digital Archivist
    - **Liz Wright**, Corporate Records Administrator
    - **Andrew Power**, Corporate Information Analyst
    - **Sue Bigelow**, Conservator
    - **Scott Redgrove,** Digital Archivist
    - **Heather Gordon**
    - **Reuben Ware**, Director
  - *Information Technology, VanMap GIS:*
    - **Jonathan Mark**, Manager
    - **Meng Li,** Chief VanMap Architect
    - **Frank DeWith,** Oracle Database Specialist
  - *Artefactual Systems Inc.:*
    - **Peter Van Garderen**, President / Consultant

- *Phase II: 2006*
  - *SDSC:*
    - **Reagan Moore**
    - **Ilya Zaslavsky**
    - **Richard Marciano**

RELEVANT OBSERVATIONS:

Ad-Hoc Committee on Archiving and Preserving Geospatial Data
Policy Advisory Network Node
GeoConnections
By David Brown et al., Library and Archives Canada

- **Introduction**
  - Part of the solution to the geospatial data preservation problem is the development of enhanced infrastructure which includes a physical infrastructure to support the collection, management, preservation, and accessibility of the information or data

- **Section 6 Technology Obsolescence**
  - Section 6.1 Data Representation
    - The majority of geospatial file formats are based upon proprietary industry standards rather than the adoption and implementation of national or ISO based standards
  - Section 6.2 Storage Technologies
    - Create copies of data on multiple storage resources

- **Section 7 Preservation Strategies**
  - Section 7.1 Temporal Management of Geographic Data
    - Creation of database versions
      - Collection of manageable states that are discrete snapshots of every geographic phenomenon that is recorded in the database
      - Allows the investigation of complicated time series spatial relationships
    - Creation of database snapshots
      - Collection of snapshots of a database at regular time intervals at static points in time
      - Allows the analysis of temporal changes in the spatial patterns of geographic phenomena

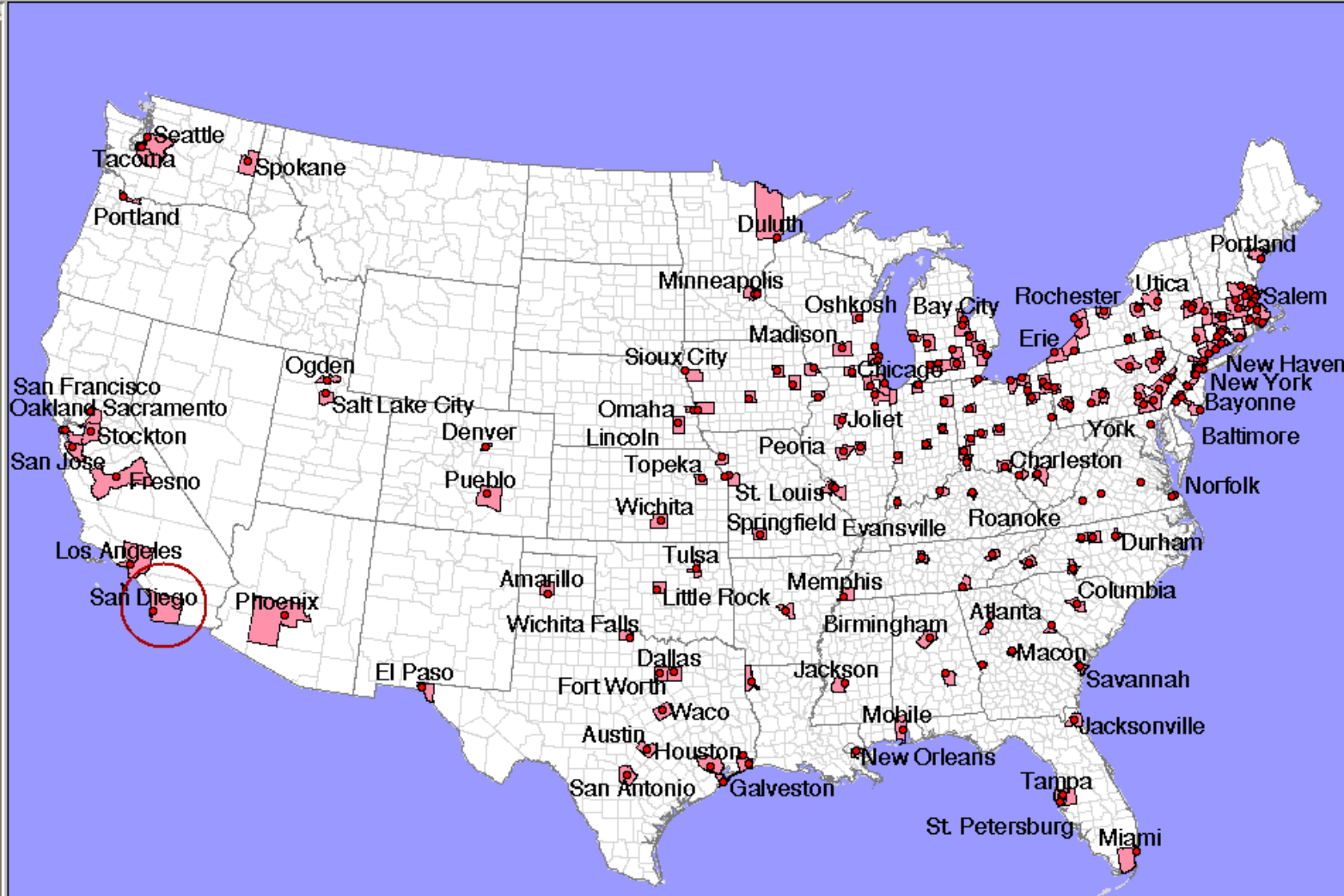# Building Infrastructure for the Preservation of Geospatial Data

- **A Few Projects:**
  - *NGDA:* *"National Geospatial Digital Archive"*
    - Partners: Stanford & UCSB
    - Funded by: Library of Congress NDIIPP
    - Objective: Create a federated network for the archiving of geospatial imagery and data

  - *GeoArchives:* *"Maine Archives of Geographic Information"*
    - Partners: State of Maine
    - Funded by: NHPRC
    - Objective: Provide long-term access to GIS records

  - *eLegacy:* *"Archival Appraisal, Accessioning, and Preservation of California's Geospatial Records"*
    - Partners: SDSC, California State Archives & CERES
    - Funded by: NHPRC
    - Objective: Develop standards for the appraisal and accessioning of geospatial data and use of data grids for long-term preservation

  - *T-RACES:* *"A Testbed for the Redlining Archives of California's Exclusionary Spaces"*
    - Partners: SDSC, UCHRI
    - Funded by: IMLS
    - Objective: Build a distributed data grid for the preservation of historical maps

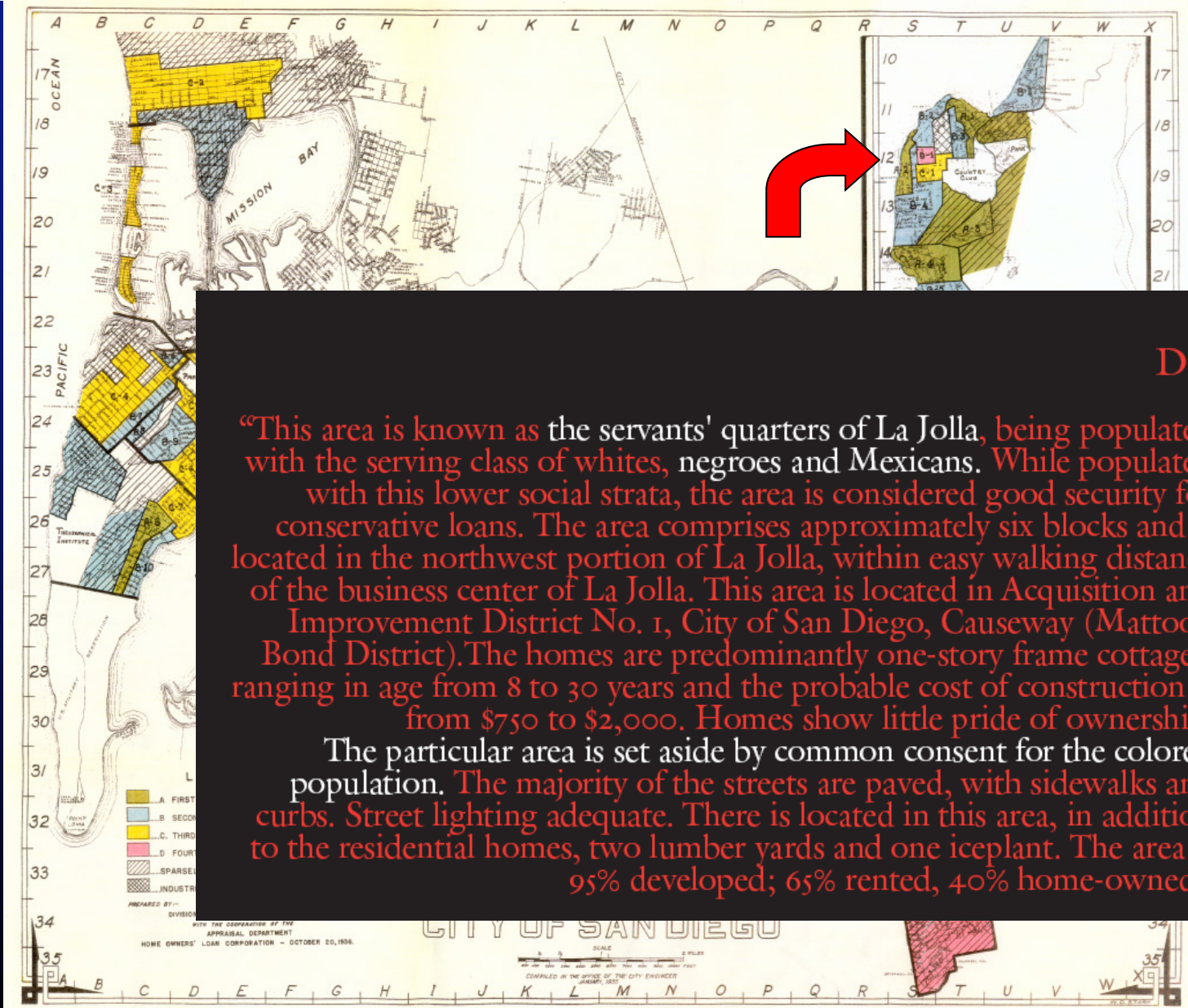UCSD

SDSC

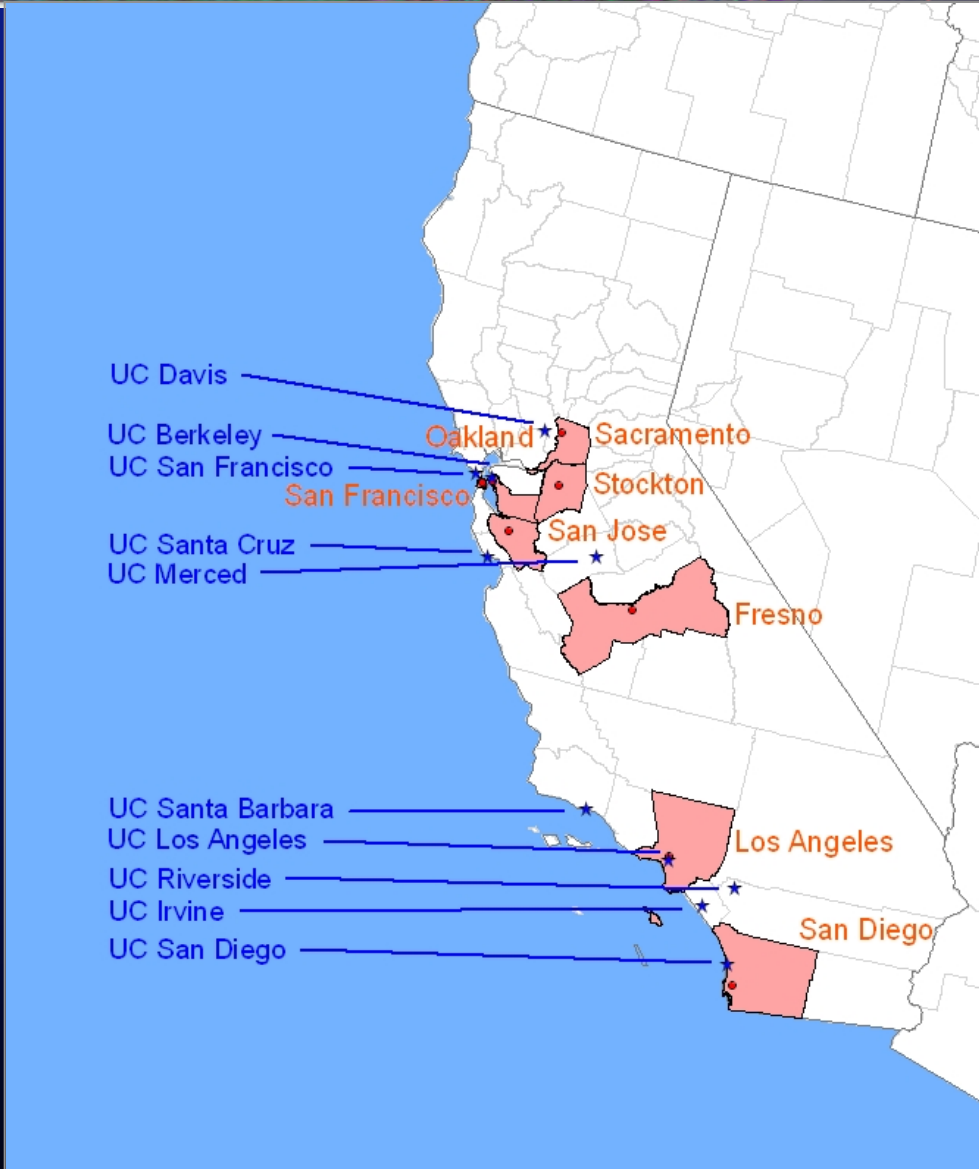# T-RACES and the Redlined Cities of America

HOLC Confidential National Security Map

D-1

"This area is known as the servants' quarters of La Jolla, being populated with the serving class of whites, negroes and Mexicans. While populated with this lower social strata, the area is considered good security for conservative loans. The area comprises approximately six blocks and is located in the northwest portion of La Jolla, within easy walking distance of the business center of La Jolla. This area is located in Acquisition and Improvement District No. 1, City of San Diego, Causeway (Mattoon Bond District). The homes are predominantly one-story frame cottages, ranging in age from 8 to 30 years and the probable cost of construction is from $750 to $2,000. Homes show little pride of ownership. The particular area is set aside by common consent for the colored population. The majority of the streets are paved, with sidewalks and curbs. Street lighting adequate. There is located in this area, in addition to the residential homes, two lumber yards and one iceplant. The area is 95% developed; 65% rented, 40% home-owned."

Mapping of UC Storage Nodes to Redlined Cities

UC Campuses	Redlined California Cities
1. UC Davis 2. UC San Francisco 3. UC Berkeley 4. UC Merced 5. UC Santa Cruz 6. UC Santa Barbara 7. UC Los Angeles 8. UC San Diego 9. UC Irvine 10. UC Riverside	1. **Sacramento** 2. **San Francisco** 3. **Oakland** 4. **Stockton** 5. **San Jose** 6. **Fresno** 7. **Los Angeles** 8. **San Diego**

Cyberinfrastructure for the Humanities and Social Sciences
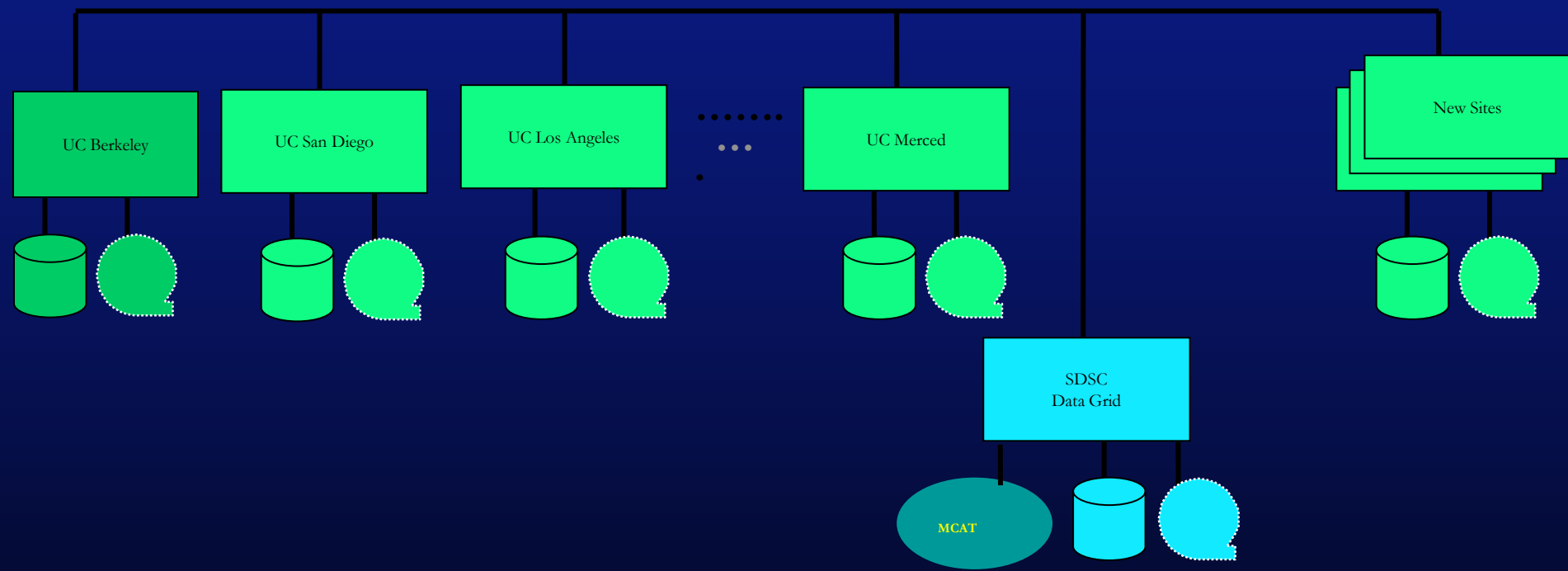
Presentation Outline

1. Introduction to GIS and VanMap
2. How VanMap is Created and Used by City of Vancouver Staff
3. VanMap: a Fluid Dynamic System
4. Is VanMap a Record?
5. VanMap as a "Potential Record"
6. What Should be Saved and Set Aside?
7. Using Data Grid Technology
8. Experiments in Migration of VanMap Data
9. On Authentic Records
10. What Data Should be Saved and When?
11. Re-instrumenting VanMap to Work with Data Grid Technology
12. On Saving Every Single Staff View
13. Creating Policies and Procedures
14. On Augmenting Data Grid Technology with Preservation Strategies

- **VanMap:**
  - Integrated Geographic Information System (GIS)
  - Streets, Fire hydrants, Utilities, Parks, Parcel boundaries
  - Information comes from City offices, County offices, State offices
  - Web-based access to city records

www.vancouver.ca/vanmap

Staff Edition

VanMap *staff edition*

Contact Us

| VanMap Home | Start VanMap | First Time Users | Announcements |

Documentation Index

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

About the Data

Layers

Join the VanMap Staff Users list

Welcome to VanMap Staff Edition v5!

VanMap now includes the proposed ward boundaries and the voting places for the October 16 vote.

VanMap is the City's answer to providing staff with quick and easy access to Vancouver's map based information. You can view, query, print, annotate, view reports, copy and even e-mail a map!
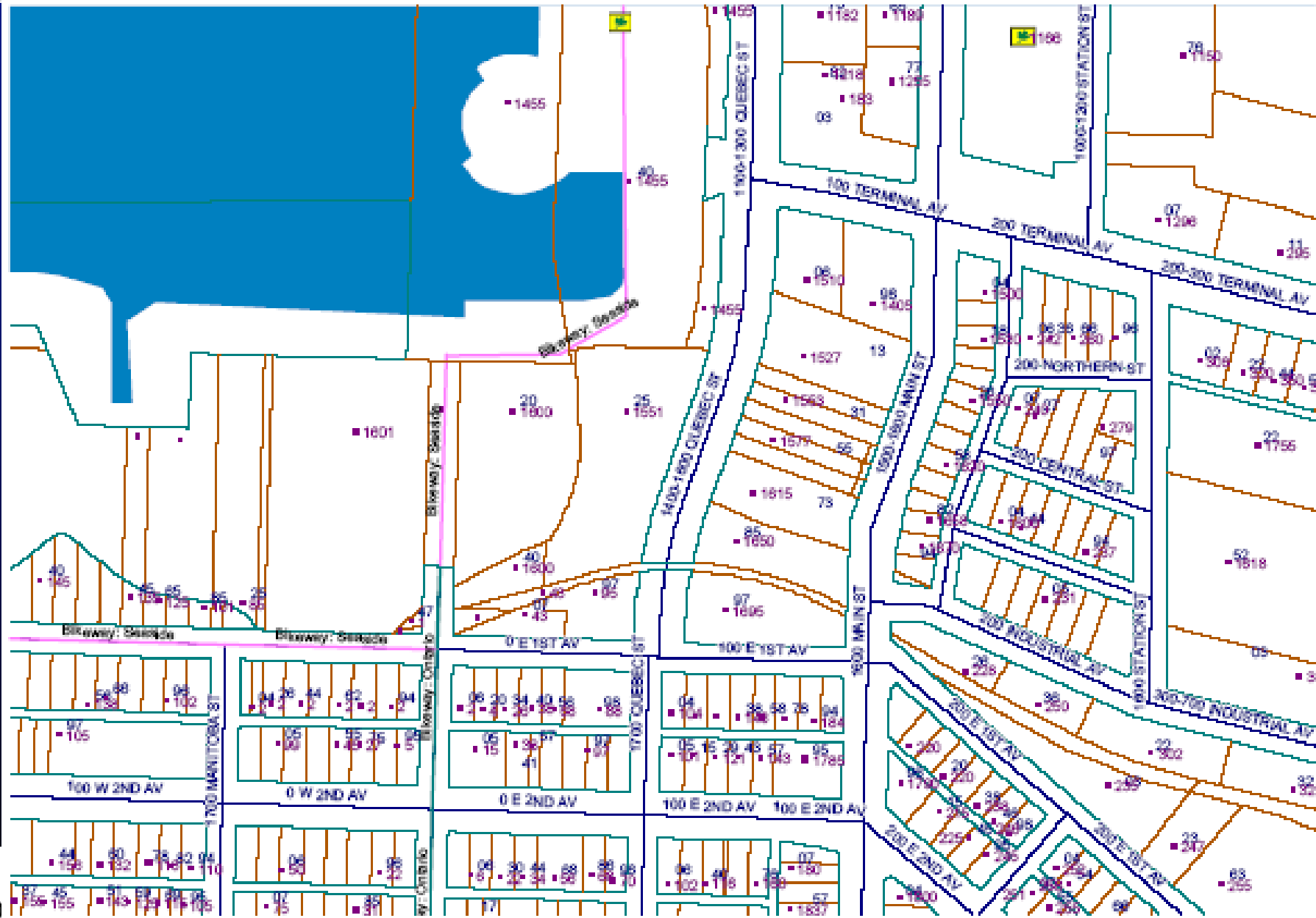
Use the alphabetical list on the left to help you use VanMap. For example, if you would like to know more about owner/address notification reports, click on R then reports or, click on O and then owner/address notification report. Everything you need is contained in this list.

VanMap is updated weekly. Click here to go to the **About the Data** page for more information or click the link in the sidebar.

Note: VanMap shows a snapshot of the City's geographic data maintained by other applications; it also includes data owned and maintained by other agencies, including City of Burnaby, First Nations, Provincial government, Federal government, or Crown agencies.
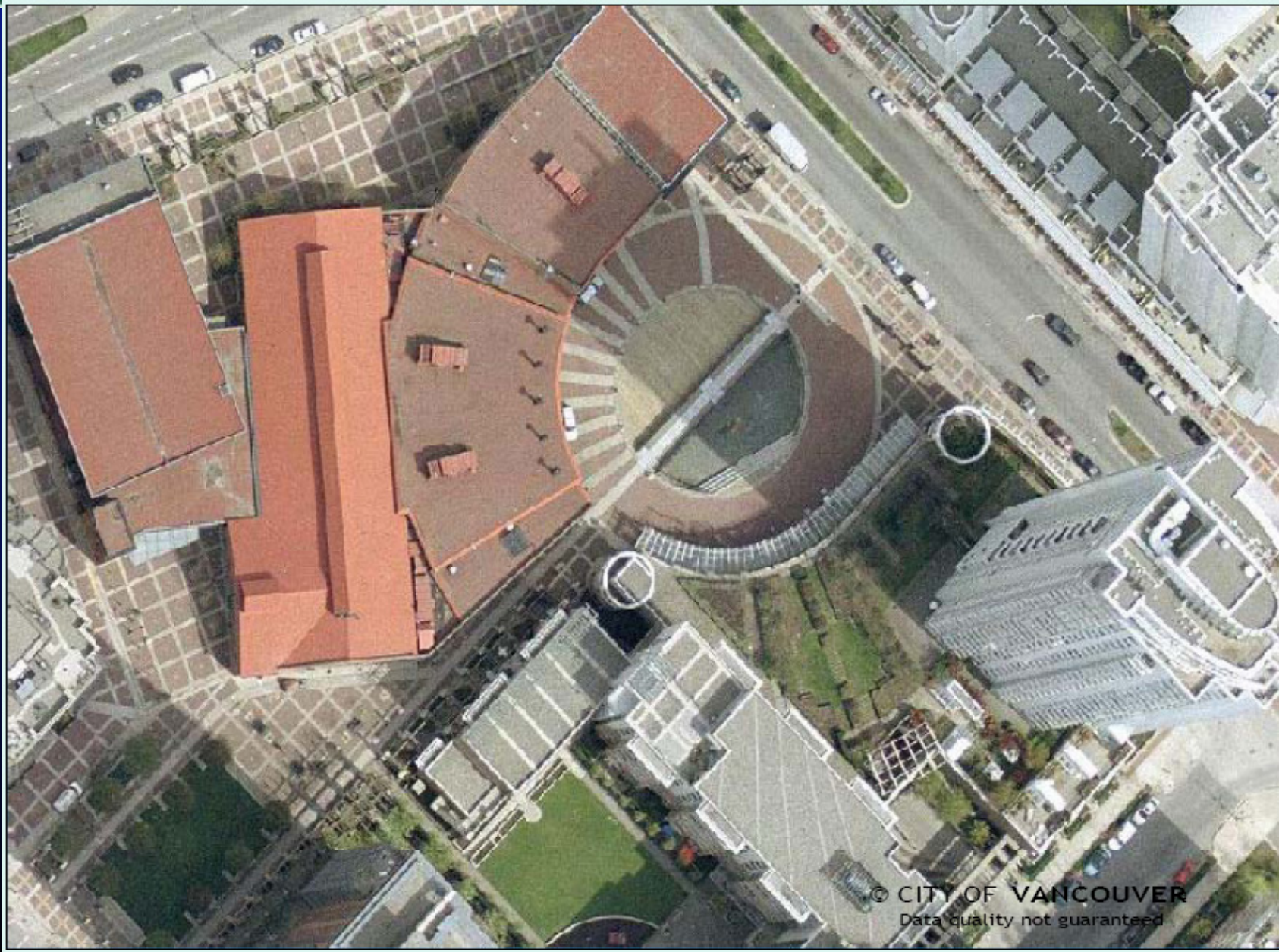
[top]

© CITY OF VANCOUVER
Data quality not guaranteed

Reports

List of Layers in Internal Version

360 Networks Duct
AT & T Cables
AT&T Manholes
BC Gas, now Terasen Gas
BC Hydro - Pole, Stub, Distribution Box, Distribution Duct, Transmission Duct, Duct Label, Manhole
Bikeways
Block Outlines
Building Footprints
Business Improvement Areas
Building Lines
Business License Sites
Cadastral Boundaries (property/lot lines)
Central Heat Steam Pipe, Manhole
City Boundary
City-Owned Properties
City Projects
Conduits, Street Lighting
Contour Lines
DCL Areas
Dedicated Fire Protection System Hydrants
Dedicated Fire Protection System Mains
Dedicated Fire Protection System Valves
Dog License Sites
Dog Off-leash Areas
Downstream Inverts
Facet Grid Layout
Facet Grid Polygons
False Creek Navigable Channel
Film Office Work Areas
Grow Operations
GT Duct
GVRD Trunk Sewers
Heritage
Hooked Sites
Junction Boxes, Street Lighting
Land Purchase Fund
Lanes
Local areas
Lot Numbers
Multiple Buildings
Non-City Streets
Non-Market Housing
Novus Duct
October 16 Voting Places
Orthophotos
Parks
Water Large Valve Labels
Water Pressure Zones
Water PRV Stations

Parks Fund
Peat Areas
Pole, Street Lighting
Pop-Growth 91-96
Private Encroachment
Private Streets
Property Addresses, Property Address Number
Property Information
Property Dimensions
Property Endowment Fund
Property Use Inspection Districts
Proposed Ward Boundaries
Protected Licenses
Public Art
Public Housing Fund
Public Places
Public Streets
Public Street Names
Reference Streets
Right-of-Way Widths
Road Closures
Service Panels, Street Lighting
Sewer Mains
Sewer Manholes
Shaw Cable Duct, Shaw Cable Pole
Shore Lines (1999) and Sea Shore Lines (2002)
Split Zones
Street Furniture
Street Intersections
Street Lighting
Street Operations District Foreman's Areas
Subdivision Categories
Survey Control Monuments
Telus Box, Duct, Pole, Stub
TeraSpan Duct
Tie Lines
Traffic Circle
Traffic Counts
Traffic Signals
Under Construction
Upcoming Projects
Upstream Inverts
View Cones
Water Bodies
Water Distribution Mains
Water Hydrants
Water Hydrant Labels
Water Large Valves
Water Transmission Mains
Youth driven organizations
Zoning Districts (with CD-1 numbers and thematic mapping)

UCSD

- **Mostly used by city departments**

- **Data maintained by departments**

- **Snapshots of data from databases published as separate layers into the VanMap GIS (128+ layers)**

- **Facilitates city decisions**

- **Effects of prior decisions can be viewed**

- **VanMap is an interactive, fluid environment display that simplifies access to current City records**
- **VanMap does not provide access to historical data**
- **The changes to VanMap are not saved**
- **The views that are created on the multiple layers are not saved and cannot be reviewed to understand the basis for prior decisions**

→ Integrating VanMap with a preservation system would enable to examine the data on which prior decisions were based

4. Is VanMap a Record?

- **Data is overwritten without being saved** *(lack of fixity)*
- **Nothing is saved and set aside for future action or reference**
  - No fixed documentary form
  - No stable content
  - No archival bond with other records
  - No record context

- **Information used to inform a government decision could be kept as a formal record of government activity**
- **This record could be archived by preserving snapshots of the city databases or by preserving the components of the VanMap system over time**
- **If the archived databases can be connected to the VanMap system, then a view can be recreated of the City of Vancouver at a prior point in time**

- **VanMap presentations can be turned into records by creating fixed representations: "setting the records aside"**

- **The impetus to preserve VanMap as a record is driven by public use of government records to support litigation, appraisal and review of prior government decisions, and as a valuable historical resource**

- **Preserving of all the city databases?**
  - Very difficult: housed at different locations, updated at different times, managed by different organizations
- **Preserving snapshots of the databases?**
  - Very difficult decision
- **Solution: "preserve a record of evidence of City actions". Options:**
  - Preserve the view represented by the composition of the VanMap layers that were displayed
  - Preserve each layer used to compose the VanMap presentation
  - Preserve the database from which the information for each layer is extracted

- **Frequency of preservation:**
  - Monthly snapshot, each time the data within the system is updated, each time the system is accessed?
- **Solution: look at the amount of time over which government decisions are made**

- **To save the changed layers and allow for queries that reassemble the layers for a given date**

- **How data grids work…**
  - Data grids are software systems that manage distributed records
  - Data grids provide uniform access mechanisms across heterogeneous storage systems
  - Data grids can manage "technology evolution"
    - Migrate data from the old technology to the new technology
  - Data grids manage the properties of the preservation environment independently of the type of physical storage resources that are used (logical file name, user namespace, storage namespace)

- **Infrastructure Independence**
  - Organize logical collections of records, independently of where the records are stored
  - Descriptive attributes (metadata) can be associated with each record *(e.g. name of the city office that provided the data, date the record was created, person who initiated the record creation, etc.)*
  - Access controls are then implemented as constraints between the logical file namespace and the logical user namespace
  - "Trust virtualization": ability to manage authentication and authorization independently of the remote storage system

- **Possible data grid usage:**
  - Each layer within VanMap can be preserved as a record in a data grid
  - By selecting which layers to compose, any desired presentation of the archived data can be assembled
  - The layers can be organized in a logical collection hierarchy by date of preservation
  - Preservation attributes are stored to identify all snapshots of desired layers that are available for a desired time period

SDSC

- **Preservation analysis scenarios**
  - Build a VanMap presentation that would have been seen by a staff member
    - By composing the desired layers for a particular point in time
  - Select a desired layer and compare two different times
    - By showing how information changes over time
  - Show the variation of the information about the City of Vancouver during a decision process
    - By comparing two different VanMap snapshots

- **Store in the data grid a "project file" which:**
  - Defines how the layers are composed
  - Defines properties of each layer (color of lines, symbols used to represent data)
  - Defines the data presented in the layer
- **A query on the system evaluates the closest previous snapshot of the layer, retrieves the GIS files, and loads the information into a public domain GIS system**
- **Preservation principles: (1) creation of an archival form of the VanMap GIS system, (2) preservation of the archival forms themselves**
  - goal = preserve information content of VanMap, even as VanMap itself evolves
  - In experiment we used: GML, Shape, Storage Resource Broker, ArcIMS

Data Exchange: received on Fri. Jan. 27, 2006

Layer Name	Group Layer Name	Feature Type	Data Source	Update
Parcel Lot	Property Info	Polygon	Oracle export Geom + tax info	daily
Orthophotos		Raster	.ecw	Every 2 to 3 years
Zoning Distr.		Polygon	SDF	weekly
City Owned Properties		Point	MS Access	weekly
City Projects	City Projects	Point	SQL Server	weekly
City Projects	City Projects	Line	SDF + MS Access (URL)	weekly
Parking Meter Rates	Traffic	SDF	SDF	weekly
Traffic Counts	Traffic	Point	MS Access: URL to repts.	yearly
Public Streets	City St. Network	Line	Oracle	One only
Web Cams		Point	MS Access + SDF	One only
Fixed Layers	Background	SDF	SDF + mwf/mwx	One only

UCSD

SDSC

Back

Search   Favorites

Address   http://cuahsi001.sdsc.edu:8080/axis/vanmapdemo.html   Go

Google   Search   19 blocked   Check   AutoLink   AutoFill

Links

VanMap Demo

Query Date

Month: May   Day: 5   Year: 2005

1995
1996
1997
1998
1999
2000
2001
2002
2003
2004
2005
2006

Submit

Query on 5/5/05

The Algorithm

1. Query the registry for specified date
2. Retrieve data grid handle
3. Download data from grid to GIS Server
4. Unzip files
5. Order / color layers automatically
6. Create .axl
7. Create a map service
8. Serve map in web browser
9. Interactive temporal interface starts

9. On Authentic Records

- **What matters is the ability to reproduce the data as interactive maps**

- **An authentic record preserves the descriptive preservation attributes and preserves the integrity of the information content**

- **The selection of the properties that must be preserved (kept invariant) needs to be defined for each GIS layer (e.g. color may not be so important)**

- **Behaviours associated with the interactive maps also need to be preserved, not so easy**

- **Frequency of preservation of VanMap data is a critical question in the design of the preservation architecture**
- **Currently VanMap manages:**
  - 1 TB of data (most of this is in the photographic images of the city)
  - 1 million files
- **If snapshots are preserved when data has changed on a daily basis, the amount of records added per year (if 100 layers were archived each week) is:**
  - 5 Gbytes
  - 30,000 files
- **Thus, preserving updates is doable. The difficulty is the creation of a standard archival form across the multiple types of databases used by government offices.**
- **A central repository could be created with snapshots of VanMap layers at regular intervals and underlying databases and backups**
- **A view could be reassembled from such a repository (not easy though)**

Storage Resource Broker Collections at SDSC (8/2/2005)	GBs of data stored	Number of files	Users with ACLs
Data Grid	Ź	Ź	Ź
NSF/ITR - National Virtual Observatory	53,862	9,536,751	100
NSF - National Partnership for Advanced Computational Infrastructure	36,149	7,539,180	380
Static collections Š Hayden planetarium	8,013	161,352	227
Pzone Š public collections	12,998	6,707,952	68
NSF/NPACI - Biology and Environmental collections	40,155	76,083	67
NSF/NPACI Š Joint Center for Structural Genomics	15,731	1,577,260	55
NSF - TeraGrid, ENZO Cosmology simulations	176,730	2,125,945	3,267
NIH - Biomedical Informatics Research Network	10,561	7,596,888	303
Digital Library	Ź	Ź	Ź
NSF/NPACI - Long Term Ecological Reserve	256	9,033	36
NSF/NPACI - Grid Portal	2,620	53,048	460
NIH - Alliance for Cell Signaling microarray data	741	84,594	21
NSF - National Science Digital Library SIO Explorer collection	2,733	1,083,998	27
NSF/ITR - Southern California Earthquake Center	131,010	2,702,421	73
Persistent Archive	Ź	Ź	Ź
NHPRC Persistent Archive Testbed (Kentucky, Ohio, Michigan, Minnesota)	100	382,186	28
UCSD Libraries archive	4,147	408,050	29
NARA- Research Prototype Persistent Archive	1,478	893,434	58
NSF - National Science Digital Library persistent archive	3,600	27,034,150	136
TOTAL	501 TB	68 million	5,335

UCSD

- **Data grids support registration of not only files, but also remote directories, databases, database tables, and SQL queries on databases. This ability to register multiple types of digital entities, makes it possible to access the multiple sources of information maintained by departments.**

- **The VanMap architecture can be extended by creating an interface that retrieves data using one of the SRB clients. VanMap can then access data from any of the storage environments or databases accessed by the SRB. This does require installing a SRB server at each storage site.**

- **Data grid technology can act as a recordkeeping system for VanMap or work in concert with a recordkeeping system**

- **The issue of saving the views of VanMap as staff use it to conduct civic business is not practical though and not addressed so far**

- E.g. a policy and procedures manual for business license applications could stipulate that certain layers must be turned on in order to conduct the procedure properly.

- We would not be reproducing exactly what staff would have seen, but what they would have been able to see when they carried out the activity

- Preservation strategy: balancing between capturing as much as we can, and capturing what will be sufficient

- What we are really proposing is a recordkeeping system. Records systems can be turned into preservation environments through the addition of preservation processes that manage assertions about authenticity and integrity.

- There is a new SDSC system that can apply preservation policies expressed as rules controlling the application of preservation micro-services *(iRODS: integrated Rule-Oriented Data System)*

UCSD

SDSC

- **The results from application of the rules are stored as persistent state information that can be subsequently queried and evaluated to prove that the management policy was enforced. Types of management policies include:**
  - Periodic check of the integrity of the records (evaluation of file checksums)
  - Periodic checks of the authenticity of the records (evaluation of the preservation metadata associated with each file)
  - Replication validation for required number of copies
  - Distribution validation for the required number of independent storage systems
  - Audit trail assessment for actions performed upon each record
  - Conformance of the system to disposition agreements
  - Conformance of the system to submission agreements
  - Conformance of standard reports
  - Migration strategies
  - Transformative migration strategy for changing file formats
  - Presentation strategy for formatting data for display

- **To enable the ability to evolve the management policies, iRODS supports three additional logical name spaces compared to the SRB technology:**
  - Logical name space for users – needed for trust virtualization
  - Logical name space for storage resources
  - Logical name space for digital entities (files)
  - Logical name space for rules
  - Logical name space for micro-services
  - Logical name space for persistent state information (preservation attributes)

Conclusion

- This case study may have broad applicability to dynamic systems in which digital entities are not saved and not set aside

- Data grid technology enables the management of distributed data collections in support of both digital libraries and preservation environments

- Data grids enforce the properties required for preservation while supporting the indexing and organizational structures managed by digital libraries

- Data grid technology support the interfaces needed to integrate distributed data under presentation systems such as GIS environments

- It is possible to build a VanMap system that accesses data from distributed repositories, accesses data from preservation environments, and supports multiple types of access and display for comparing current and historical records

For More Information

Richard Marciano

marciano@sdsc.edu