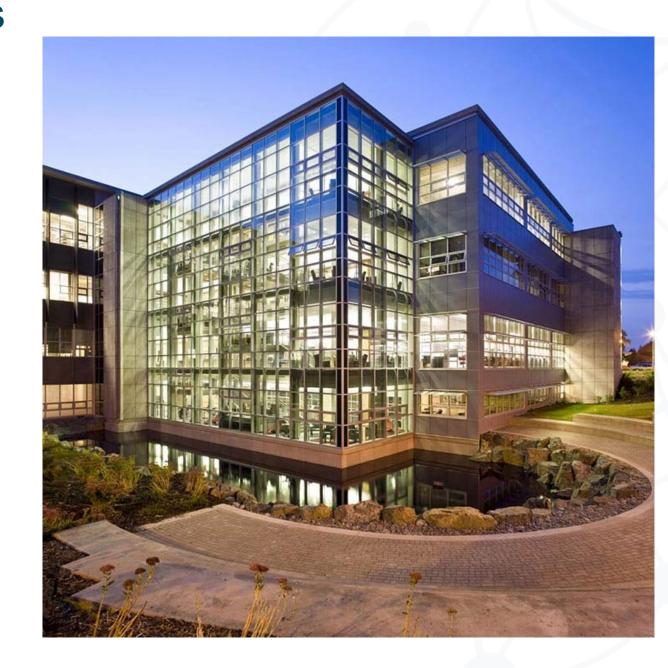


# Petabyte-scale Challenges for Research Library Collections

Corey Davis
Digital Preservation Librarian
University of Victoria

## University of Victoria Libraries

- 140 employees
- 2 million physical resources
- 162,000 journals and 850 research databases
- 2,283 linear metres of manuscripts and archival materials







What do our digital collections look like?

## Digital collections at research libraries

- 1. Materials licensed from publishers and made accessible via the library
- 2. Unique or distinctive local collections managed by the library





Licensed materials

## Scope

- Mostly journal articles and books
- Annual license fees (i.e. we don't own the materials)
- Discovery via library management systems and tools like Google Scholar
- Access via publisher's website
- Long-term preservation via external organizations like Portico, CLOCKSS, and Scholars Portal





## What's involved?

- Licensing
- Budgeting (\$USD + inflation = headaches)
- Managing title lists
- Creating and maintaining metadata records
- Subscribing to preservation services to ensure post-cancelation access
- Open access advocacy, managing author funds, research help, etc.





Distinctive collections

## Scope

- Research data
- University publications (journal articles, pre-prints, theses and dissertations, etc.)
- Materials digitized from the analog collection (maps, newspapers, sound recordings and film)
- Born-digital records (from the university and from donors)
- Web archives
- Open educational resources (OER)



## Ocean Networks Canada

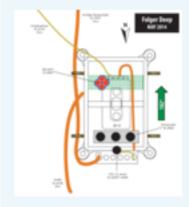
Discover the ocean. Understand the planet.

(Ocean Networks Canada)

Borealis > University of Victoria Dataverse Collection > Ocean Networks Canada >

# Six years (2009-2015) of oceanographic temperature, salinity, pressure, density and dissolved oxygen observations from a Vancouver Island shelf cabled observatory.

Version 1.1



Ocean Networks Canada, 2020, "Six years (2009-2015) of oceanographic temperature, salinity, pressure, density and dissolved oxygen observations from a Vancouver Island shelf cabled observatory.", https://doi.org/10.5683/SP2/5AMBHV, Borealis, V1,

UNF:6:nWcbYa+XduvV+k6o8Cl6HQ== [fileUNF]

Cite Dataset ▼

Learn about Data Citation Standards.

Access Dataset →	
Contact Owner	Share
Dataset Metrics	
263 Downloads 🕢	

# **UVicSpace**

<u>UVicSpace Home</u> → <u>Faculty of Graduate Studies</u> → <u>ETD (Electronic Theses and Dissertations)</u> → View Item



A 100-year retrospective and current carbon budget analysis for the Sooke Lake Watershed: Investigating the watershed-scale carbon implications of disturbance in the Capital Regional District's water supply lands

Smiley, Byron

URI: http://hdl.handle.net/1828/6117

Date: 2015-05-01

#### Abstract:

Northern forest ecosystems play an important role in global carbon (C) cycling and are considered to be a net C sink for atmospheric C (IPCC, 2007; Pan, et al., 2011). Reservoir creation is a common cause of deforestation and when coupled with persistent harvest activity that occurs in forest ecosystems, these disturbance events can significantly affect the C budget of a watershed. To understand the effects of these factors on carbon cycling at a landscape level, an examination of forest harvest and reservoir creation was carried out in the watershed of the Sooke Lake Reservoir, the primary water supply for the Greater Victoria area in British Columbia. Covering the period between 1910 and 2012, a detailed disturbance and forest cover dataset was generated for the Sooke Lake Watershed (SLW) and used as input into a spatially-explicit version of the Carbon Budget Model of the Canadian Forest Sector 3 (CBM-CFS3). The model was modified to include export of C out of the forest system in the form of dissolved organic C (DOC) into streams. The fraction of decaying C exported through this mechanism was tuned in the model using DOC measurements from three catchments within the SLW. Site-specific growth and yield curves were also generated for watershed forest stand types, in part, by using LiDAR-derived site indices. C transfers associated with disturbances were adjusted to reflect the disturbance types that occurred during the 100-year study period. Due to the removal of C resulting from wildfire, logging and residue burning, as well as deforestation disturbances, total ecosystem C stocks dropped from 700 metric toppes of C per bectare (tC ba-1) in 1910 to their current (2012) level of ~550 tC ba-1

Search UVicSpace	
Search UVicSpace This Collection	Go

#### Browse

All of UVicSpace

Communities & Collections

Authors

<u>Titles</u>

Subjects

Supervisors

By Submit Date

This Collection

Authors

Titles

Subjects

Supervisors

By Submit Date

#### My Account

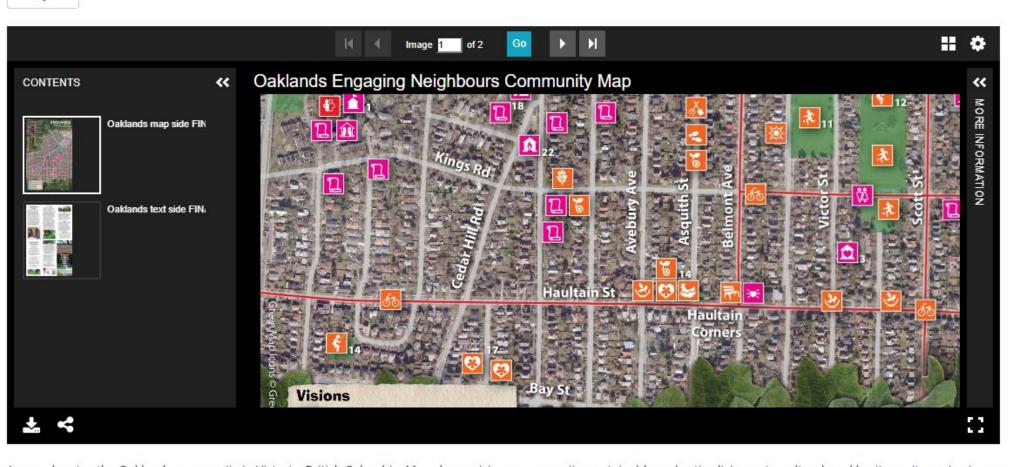
<u>Login</u> <u>Register</u>

#### **Statistics**

View Usage Statistics View Beta Statistics Home

## Oaklands Engaging Neighbours Community Map

Analytics



A map showing the Oaklands community in Victoria, British Columbia. Map shows visions, community, sustainable and active living, arts, cultural, and heritage sites using icons from the Green Map System. Reverse of map features information about Oaklands; assets, visions and actions; and information on the Community Tables: Engaging Neighbours initiative and its goals. Created in conjunction with the United Way of Greater Victoria, BC Healthy Communities, the University of Victoria Department of Geography and the University of Victoria's Office of Community-Based Research. This map is part of the green map system.





Realistic TR-284. The newest electronics! The finest styling! i stacen that better than Her sutometic 2-channel/4-channel sensing, automatic and 4 channels manual track selection; volume, tone, stereo belance controls. Player activates when cartridge is inserted for instant music -ack at four turns off when certridge is removed. Program indicator lights esence never and illuminated "4-channel" indicator, 15 x 5 x 11 3/4", eme to make Each speaker 7 7/8 x 5 1/2 x 9 7/8". AC es to call it



GHTS 609 YATES STREET Phil Smith, Manager ager 384-6411

STORE OR WRITE TO: BOX 340

had a fee increase for three years, Anderson says it represents an annual increase of 2.7 per cent per annum Stockers Tie which I don't think even keeps up with increases in office expenses."

The result of the referendum, amornoed Thursday is an endorsement of BCMA polnew fee achedule, says Ander-

#### Find Offers Theft Lead

NEW YORK (AP) - Detectives recovered \$7,600 in stolen securities Tuesday and called it an important lead in Their investigation of the emonthly-executed weekend helat of ninre than \$1 inition from the fashionable Pierre

Police said the securities were taken from the Drake Hotel on Christmas in a \$15,000 robbery authorities fur-Here was pulled off by the some holdup team.

# Nanaimo Club

Stockers North Americans by that, effective Jun. 1, don- pulled goalkeeper Bob Dobiah tors start billing under the and Gerry Caldwell scored with 28 aeconds remutaing to give the Stuffy McGinnis Hor-"The conients of the fee key League leaders a 5-5 tie with Namimo C. and R. Halievers, of the Pacific Coast America Hockey League in an exhibition game Thursday night at G. R. Pearles Arena.

Errol Goulden had three govis for Stockers, who raced to a 4-b lead by the middle of the second period. Dave Wilhams got the other Stockers' goal, iter. Woodward, Don. Brown, Larry Limacher, Joe. Piper and Dave Jones second for Nanaimo.

### NOMONE TILL MARCH

HONDA NEXT HER LINN

PEARSON

## Take Off Fat With **Home Recipe Plan**

cover and article past estain the little hard go to your drogstors and sake for laran. Fore the info spin boats and and enough transferred ports to fill the boats. Take two labelesprometri twice at the plan and help bring back along as signified and follow the stenders. Note how against the plan and help bring back along a significant and follows: Biaran Hadauing Ylen.

ose bulky fat and nely regain youthful appearing and active.

It's slapple how quickly one slender more gracuful carves; if may loss pounds of smaightly fat reducate pounds and inches of right in your own home. Make excess fat don't disoppose from this home racip, yourself. It's mack, thin, arms, abdomen, hips, easy, no temities of all and costs culvus and ankles ) as extern the stenderness. Note how quickly If your first purchase does not blogt disappears how much dine you a simple easy way to batter you lee! More size,

Bills for patients who have been sent on to a specialist would not be rejected for some time because the commission in such cases pays only when treatment has been completed or involves an ciaborate orocedure or a setof tests that is well under

CLEAN CLAY FILL Hauling charge only. also sand and gravet. 478-5613 or 477-4142



and highway driving ease.

TEST DRIVE TODAY

\*2318

F.O.B. VICTORIA

PETER POLLE

YATES AT COOK

384-1144

hide

Lytton

NEWS

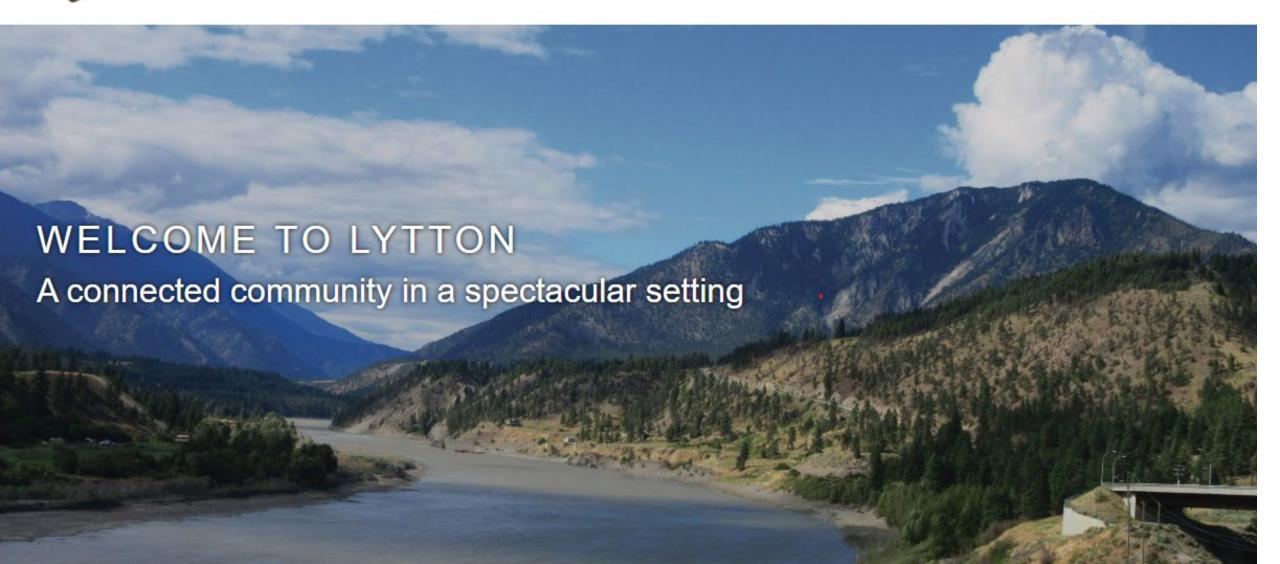
TOWN

LOCAL GOVERNMENT

LIVING HERE

VISITING

HOW DO 1?

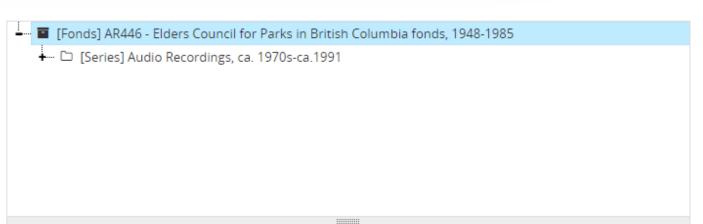


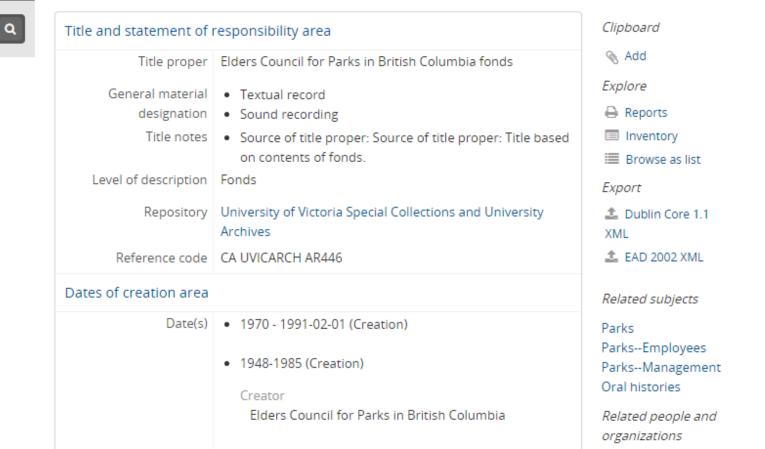


Ouick search

Search

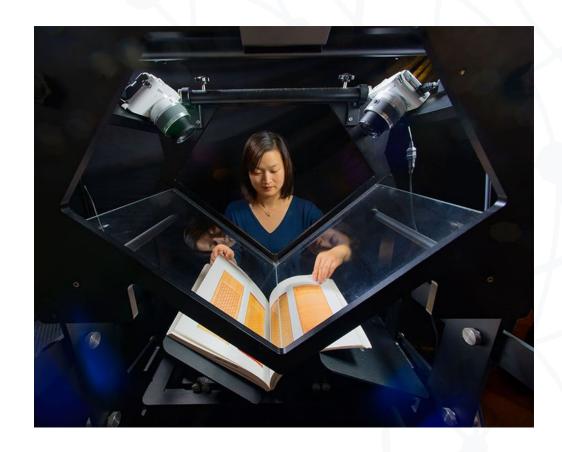
## Fonds AR446 - Elders Council for Parks in British Columbia fonds





## What's involved?

- Born-digital accessioning and large-scale digitization
- Metadata creation and persistent identifiers
- Managing repository and access platforms
- Long-term preservation processing
- Storage infrastructure







Key challenges with our distinctive collections



## Here are some synonyms for "haphazard":

- 1. Random
- 2. Casual
- 3. Arbitrary
- 4. Chance
- 5. Unplanned
- 6. Hit-or-miss
- 7. Unsystematic
- 8. Disorganized
- 9. Unscheduled
- 10. Accidental
- 11. Indiscriminate
- 12. Sloppy
- 13. Chaotic

S Regenerate response



7

Send a message...



# Workflows, organizational structure, and capacity

- Organizational silos
- Knowledge sharing issues
- Communications issues
- Expertise is not widespread enough



## Digital asset management

- Diversity of materials and platforms
- Hard to get a comprehensive, high-level overview of all digital assets
- This means its hard to plan and be strategic



## Growth

- 100 Tb of managed data and growing quickly
- A handful of new accessions can quickly push us towards petabyte scale





Growing up

## Open Archival Information System (OAIS) model

- Initially developed by the space data community
- Preservation must be disentangled from access copies
- AIPs are platform-independent and contain preservation metadata





## Archival Information Packages (AIPs)

- BagIt standard for packaging
- PREMIS metadata for integrity and authenticity
- Ideal is to have multiple georedundant copies using different underlying technologies
- Restricted and sensitive data is a particular challenge



## Remote storage

- OpenStack Swift (Ontario Libraries Research Cloud)
  - Three copies at campus data centres in Ontario
- LOCKSS (WestVault)
  - 5 copies at campus data centres in Western Canada
- AWS (S3 Glacier)
  - Copies of last resort across multiple availability zones, both in Canada and internationally



## Issues with remote storage

- PIA / security
- Scalability / organizational viability
- Intelligent tiering
- Costs / Lock-in





## Local storage

- Archivematica used to create AIPs
- AIPs stored on 15tb network shares backed up to Kamloops via BCNET Educloud Backup Service



## Local storage issues

- We're quickly outgrowing this 'business operational' approach to storage
- Working with Research Computing Services on the possibility of using OpenStack
- We need 'intelligent tiering' to keep costs low
- Need to combine best of onsite and offsite for restricted/sensitive data





What's we're doing

## Digital Collections Growth Strategy Task Group

- How do we optimize workflows and organizational structures to accommodate anticipated future growth in our unique digital collections?
- What technologies do we need to effectively manage this growth?
  - DAMS
  - Storage
- What expertise and staffing levels are required to effectively manage this growth?



## Let's talk storage!

- What would you do?
- Lots of stuff, and growing quickly
- Need to scale sustainably and avoid lock-in or proprietary solutions
- Some data is restricted/sensitive
- Need multiple copies offsite
- AIPs do not need to be instantly available





Thanks! coreyd@uvic.ca