The NEPTUNE Canada Cabled Observatory Data Management System

Capturing and Delivering Terabytes of Data each Day
NEPTUNE Canada Quick Facts

- **NEPTUNE Canada is a regional cabled observatory**
  - 800 km cable loop from Port Alberni (BC)
  - Built and operated by the University of Victoria
  - Represents a total of $82M investment
  - 5 nodes with about 120 instruments, hundreds of sensors
  - Cable Installation in June, nodes and instruments: '08
- **VENUS is a “prototype” for NEPTUNE**
  - 3 km + 30 km lines, ~ 30 instruments, similar concept
NEPTUNE Canada Components

Benoît Pirenne, NEPTUNE Canada, UVic, Victoria, BC

- Project structure reflects components to be built:
  - Undersea infrastructure: engineering (Peter Phibbs, Alcatel Submarine Networks)
  - Instrumentation: science (Mairi Best, with PI teams)
  - Software: IT (Benoît Pirenne, in house development)
  - Education & Outreach: unfunded
NEPTUNE Canada: underwater plant

Benoît Pirenne, NEPTUNE Canada, UVic, Victoria, BC
NEPTUNE Canada: Underwater Plant
- **Instruments**
  - **VENUS**: 3 nodes, ~30 instruments, ~120 sensors
  - **NEPTUNE**: 5 nodes, ~120 instruments, ~400 sensors
**Data Management and Archiving System (DMAS)**

- Three components:
  - Data acquisition & storage
  - Data access
  - Instrument and Infrastructure control and monitoring

- Features:
  - Very fast data delivery to users
  - Real-time event detection capabilities
  - Event reaction capabilities
  - Fast access to the archive
Challenges:

- Large variety of instruments
- Three types of data flows
  - Scalar (temperature, pressure, amplitude, ...)
  - Multi-dimensional (ADCP matrices, images, ...)
  - Uninterrupted streams (acoustic data from hydrophones, video)
- Wide variety of data formats, instrument protocols
  - ASCII, binary, push, pull, ...
- Enormous differences in data rates
  - CTD: bits/sec $\leftrightarrow$ HDTV camera: Gbps
NEPTUNE Canada Data Acquisition

Benoît Pirenne, NEPTUNE Canada, UVic, Victoria, BC

- **Challenges:**
  - Event detection
    - “if $sensor > threshold then do this”
    - “if {starfish} in $cam-video-stream then $take-cam-hires-still”
    - “if $hydrophone-stream contains whale with probability > 80%”
  - Then ...
  - Policies for event reaction
    - Reaction can be email
    - Reaction can be use of shared instrument
    - Must have ranking/priority per user on the use of shared instruments
Challenges:

- Storage and data compression choices
  - Current estimates for data flow:
    - dominated by HDTV (~1 Gbps native)
    - 1 HDTV at Folger Passage at 8 hrs/day ==> ~25TB/day!
  - Can we afford to store that?
  - Will anybody need it later?
  - How much will we loose if we compress?
  - Compression methods are many and improving. Hoping to decrease volume by a factor 100
- Metadata collection/value-added to data through pre-processing
NEPTUNE Canada Data Acquisition

Benoît Pirenne, NEPTUNE Canada, UVic, Victoria, BC

- **Expectations:**
  - Current estimates (with compression) call for about 47 TB/year
  - + at least one safe copy elsewhere
  - ==> ~100TB/year for the first 3 years or so.
  - Technology available today, but total cost of ownership must be controlled to limit operations costs
  - Local copy? Outsourced storage to Amazon, Google, UCSD?

- **Data distribution:**
  - Download volumes by users unknown
  - Looking at limiting downloads and providing data centre-based pre-processing, visualization.
Data Acquisition

Benoît Pirenne, NEPTUNE Canada, UVic, Victoria, BC
Data Access

Data Access Methods

- Interactive (web search forms)
- Computer-to-computer access (web services)
- Possibility for PIs to host event detection modules at the shore station, data centre or at home
- Future: social network with many web-based data processing and visualization functions

Data Formats

- Scalar sensors: we'll offer what users want. So far for VENUS: CSV, Matlab. Coming up: NetCDF.
- Other complex data: JPG, mp3, ...
- Tried xml with ADCP data: got in trouble with users...
Conclusions

Benoît Pirenne, NEPTUNE Canada, UVic, Victoria, BC

- NEPTUNE data are coming!
- VENUS has been an invaluable learning tool for us
- Prospects of real-time data management are exciting (event detection and reaction)
- Challenges of the exploitation of so much good data coming so fast to everyone are going to be:
  - exciting scientists
  - will quickly generate new needs on the supporting software systems