

De-portal-ment

the manner
in which one conducts
a portal

Chris Van Holmes
DART Coordinator

Development

Development must address these issues/concerns

- Users are diverse with many different purposes
 - Regional research scientists, Energy supply modelers, Financial institutions, Congress, Sports fishermen, State Agencies, Tribes, K-20
- Project development must fit into funding limitations
 - Success means meeting project goals within budget and time constraints

Multi Year Effort Large Budget

C B E O

Chesapeake Bay Environmental Observatory



demonstrating the transformative power
of **cyberinfrastructure**
for environmental science and engineering

- **CBEO Participants**
- [National Science Foundation](#)
- [Chesapeake Community Modeling Program](#)
- [University of Delaware](#)
- [Drexel University](#)
- [Hampton University](#)
- [Johns Hopkins University](#)
- [San Diego Supercomputer Center](#)
- [University of Maryland Center for Environmental Science](#)

Chesapeake Research Consortium's 5 year NSF Award:

\$2.15 million

Goals for CBEO

- Develop an online prototype of integrated datasets.
- Measure and model Chesapeake Bay hypoxia
- Provide outreach and education on Chesapeake Bay environmental issues

CBEO

Chesapeake Bay Environmental Observatory



demonstrating the transformative power
of **cyberinfrastructure**
for environmental science and engineering

Welcome,
David
Jasinski

Logout

- Search
- myWorkbench
- Contribute
- GEONtools
- CBEOtools
- UserProfile
- Collaborate
- Grid
- Portal Use

Home **DASH** Hydroseek R-Interpolation

DASH

Navigation toolbar: +, -, Home, Info, Print, Full Screen, Back, Forward, Home, Search, CIMS, Register, ESRI, ESRI Support Center, Help, CUAHSI

- CIMS
- NWIS_DailyValue
- EPA
- NWIS_IID

Results

- HIS (-76.3992, 38.4804)
 - CIMS
- BaseMap (-76.3992, 38.4804)

HISID Results

1 feature at (-76.278, 38.206)

[CIMS](#)

Navigation

Overview

Small Scale Time Limited

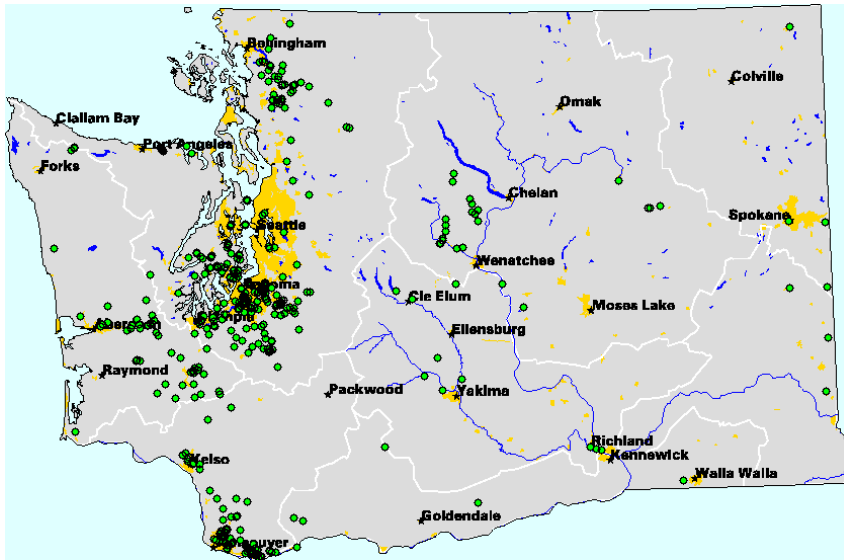


Washington
NatureMapping Program

- A central place to keep volunteer monitoring and restoration data.
- Extend water quality and biodiversity coverage in Washington.

Water Biodiversity Module Data Bank

- Very limited development time
- One time budget of \$10,000
- Currently 457 monitoring and restorations sites from May 1989- Oct 2009.



Development Styles

Wikipedia list 56 software development philosophies with
the caveat:

“This list is incomplete”

From “Agile Software Development”

thru “KISS”

to “Zero One or Infinity”

Find the best compromise between :

– Complete Upfront design

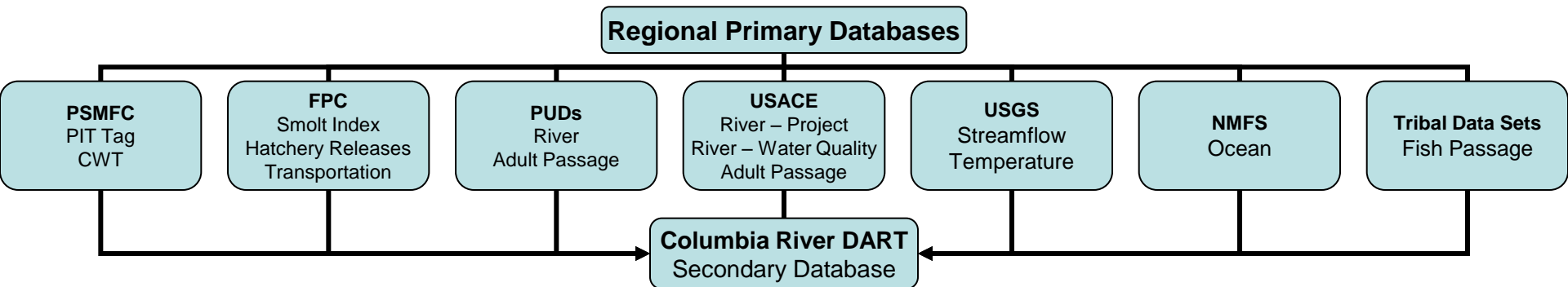
- » Implementation issues fixed before launch
- » Complete product envisioned
- » Adjustments to changing requirements can be costly

– Rapid Prototyping

- » Responsive to changing requirements and scope
- » Bugs require re-development
- » Product delivered to users quickly
- » Include development & feedback cycle

What are you Porting?

- Identify crucial sources
- Forge personal relationships
- Integrate temporal and spatial data



Personal Relationships Provide:

- Improved Access
 - Non-anonymous user of primary source
- Improved data quality
 - Direct feedback loops
 - Suggestions and insight
- Provides point of contact
 - changing data needs
 - identifying and resolving transfer issues

Acquire Data

- ODBC
 - Direct access to remote database
 - The DBA's Dream
- FTP
 - User and anonymous access to remote files on server
 - Work with primary source to create formats and content
- Email
 - Friendly but cumbersome
- HTTP
 - Automated scripts to query primary online sources
 - DART's most common method
- Manual Data Entry
 - Necessary evil for some datasets i.e. Historical adult passage

Monitor loads and sources

Automated Scripts and Processes to

- Check retrieval of data
- Parse to standardize formats
- Monitor for unscheduled format changes
- Monitor data loads into DB
- Delete duplicate data rows
- Filter data for aggregation

Synchronize to Primary Sources

- Changes in primary source data must be mirrored.
 - Use regular loads to provide synchronicity
 - Small data sets can be updated with each load
 - Large datasets can be updated periodically using batch processes.
 - Impacts user access

Data Delivery

- Deliver Data in usable form
 - Deliver format to simplest common denominator
 - Ensures usability to widest possible audience.
 - Provide alternative formats
 - Simple delimited tables to publication quality presentations
 - Make machine accessible
 - Reduces user costs for data access.
 - Allow repeat users to automate queries
 - Plan on unexpected users and uses
 - Financial investment companies are a significant user base for DART
- Balance best practices with delivery speed
 - Waiting users will become frustrated

Quality Assurance

- Know the nature of the primary sources
 - Acquire and deliver metadata
 - Nulls and codes
 - Appropriate ranges and values
- Filter data before aggregation
 - Eliminate outliers and errors
- Provide feedback to primary sources
 - This step improves the quality of the data as well as the relationship between the source and the portal

Security and logistics

- Know your vulnerabilities
 - SQL injection
 - Denial of service
- Make sure your backup plan works
- Balance
 - Control user access to safeguard public use
 - Provide flexible query choices to maximize user base
- Schedule data loads and maintenance to reduce impact on users

Identify user patterns to improve product

- Identify broad usage patterns
 - Data load and backup timing
 - Balance availability and reliability
- Satisfy targeted users
 - Leads to product development and refinement

