Second-Tier Database for Ecosystem Focus

Annual Report
1999 - 2000

DOE/BP-00004124-1

November 2000

This Document should be cited as follows:

Anderson, James, "Second-Tier Database for Ecosystem Focus", 1999-2000 Annual Report, Project No. 199601900, 20 electronic pages, (BPA Report DOE/BP-00004124-1)

Bonneville Power Administration
P.O. Box 3621
Portland, OR 97208

This report was funded by the Bonneville Power Administration (BPA), U.S. Department of Energy, as part of BPA's program to protect, mitigate, and enhance fish and wildlife affected by the development and operation of hydroelectric facilities on the Columbia River and its tributaries. The views in this report are the author's and do not necessarily represent the views of BPA.

Second-Tier Database for Ecosystem Focus

Annual Report

October 1, 1999 - September 30, 2000

Contract number
00004124

BPA Project Number
96-019

Prepared by
Chris Van Holmes
and Christine Muongchanh
University of Washington

October, 2000

Table of Contents

Abstract

The Second-Tier Database for Ecosystem Focus (Contract 19601900) provides direct and timely public access to Columbia Basin environmental, operational, fishery and riverine data resources for federal, state, public and private entities.

The Second-Tier Database known as Data Access in Realtime (DART) does not duplicate services provided by other government entities in the region. Rather, it integrates public data for effective access, consideration and application.

Introduction

Information resources for the Columbia Basin are resident among a diverse number of Governmental organizations within the region. Effective management and research of river and fishery issues frequently require timely data access in a format useful for comparison, analysis and decision support. It is precisely this requirement that DART addresses.

DART (Data Access in Realtime), is a second-tier database warehouse containing historic data back as far as 1962, and current real-time data maintained by daily accesses to regional primary-tier sites. By design, DART has been developed to integrate diverse datasets and allow data be displayed in text and graphical formats via an interactive forms based web-site (http://www.cqs.washington.edu/dart/). Project personnel work closely with the primary-tier organizations to ensure quality and timeliness of the data.

Project Activity

Objective 1 -- Support BPA and ESA Activities

Provide monitoring and evaluation

Regional integrated electronic products and services were distributed via the Internet throughout the fiscal year to support BPA and ESA mandated activities. Real time interactive services supported the BPA FWP and the Technical Management Team in adaptive management and recovery efforts. Products to support ISAB were developed

Provide electronic data integration

Data acquisition from numerous primary sites with the region was performed on a daily basis. Datasets, products and services were integrated within a relational

database based on date and location, then immediately made available over the World Wide Web and Internet. Text and graphical formats provided current real time information as well as historic data as far back as 1962.

Objective 2 -- Provide Public Services

DART provides a public Internet interface year-round to support planning and decision making for the operation of the Federal Columbia River Power System. Active users include Federal, State and interested public and private entities. These services compliment and enhance the value of primary data sites and are an important step toward a regionally integrated data services.

DART Coordination

Inter-organizational

Frequent contacts were made with PSMFC, U.S. Army Corps of Engineers, Grant County PUD, USGS, SnowTel, PFEL, NOAA, and Fish Passage Center to coordinate data transfer and acquisition concerns for the calendar year 1999 cleanup, for calender year 2000 setup, and for daily operations. Specifications relating to system requirements and query modifications were coordinated with BPA FWP. Anomalous wild steelhead counts at mainstem dams were corrected with USACE assistance. USACE was advised on data quality control and error-checking of adult passage data. Development of a procedure to update previous years PTAGIS datasets was initiated with PSMFC. Meta-data issues were clarified with USACE and SnowTel. The Technical Management Team was supported throughout the migration season. Reports, assessments, and new database tools were provided to ISRP and ISAB. We also provided assistance to NASA in adding DART to their Earth Science and Global Change Master Directory.

In-House Support

In-house support was provided to Dr. John Skalski's group, including datasets and application support throughout the year.

Historical Data Set Development

The Coded Wire Database, Version 3.1 was installed on the DART warehouse system in February 2000. Reporting capabilities, views, screens and loading programs were developed to integrate the complete CWT database into the latest version of DART. Although the data set was historic dating to 1997, all templates are now in place to load a more recent version as soon as arrangements can be made with PSMFC.

A historical life-cycle dataset containing redd counts, spawners, and recruits was developed.

Acquisition of historical temperature and precipitation data from NOAA, USBR, SnowTel, and USGS sites to augment the daily data continued. Monthly updates of the Coastal Upwelling and Moored Buoy datasets from Pacific Fisheries Environmental Laboratory were automated. The acquisition of historical flow and fish passage datasets from 1934 on was initiated.

Daily Services and Operations

Daily Data Acquisition

Hourly riverine values including outflow, spill, water temperature and a variety of water quality values were loaded, updated and processed into daily values. Water quality data is loaded several times each day to provide real-time pictures of river conditions. Final tally of riverine records exceeded 450,000 hourly data rows with 19,000 daily row averages calculated as of October 1, 2000.

Tagging session header files and PTAGIS observation records were acquired from PSMFC on a daily basis. The current year's observation total exceeded 690,000 valid records. Release and observation site datasets were updated to reflect changes and additions.

Daily flow, stage and intermittent temperature data is acquired from 10 USGS sites to provide information concerning headwater conditions. Daily temperature and precipitation records from 58 sites throughout the region provide information on weather and climate. Adult passage and smolt passage indices were loaded daily from 12 mainstem dams.

As of 2000 fiscal year-end, DART database has grown to 13.5 gigabytes.

DART Maintenance

Pre-migration date and specification changes were implemented and a variety of program edits and modifications were completed in preparation for the start of the migration season.

Normal DART maintenance was conducted throughout the year. Data loads were monitored and reviewed on a daily basis. Daily backups, performance tuning, indexing rebuilds and monitoring of incoming web user database access continued. Year-to-date updates were performed daily for adult passage counts and weekly for PTAGIS data. With cooperation by PSMFC, the weekly PTAGIS YTD update was automated. Occasional updates of water quality and smolt index datasets were performed as needed.

Quality Control

Data quality was monitored throughout the contract on a daily basis and adjustments made as needed. Primary sites were contacted concerning data omissions, anomalies and format changes.

Reloading of datasets from primary data sites occasionally re-introduced previously corrected errors. These datasets were then re-edited.

Water quality, stream flow, adult passage and smolt index data loads were modified as needed to accommodate occasional changes in primary site source files.

Loads, reports, and queries were updated to eliminate duplicate data rows.

DART Web Support

Web support involved helping external users upgrade their daily scripts used to query DART (not through our established web page interfaces). Data and meta-data requests from the public were filled as they were received. Issues with primary data raised by users were resolved.

Reports and database objects were modified to accommodate user requests when appropriate. The DART file nomenclatures were updated to eliminate a browser caching conflict, allowing users unencumbered access to the current data.

A map of the major data collection locations contained in DART was added to the web site. The map includes information on the type of data and the range of data collected at each of the major sites.

A test version of javaDART which runs as a Java application on any client system with the Java 2 Run-time Environment was developed and deployed. This version displays data from the DART database in line graphs, histograms, and text listings. We also purchased a database connection server (JSVR from Caribou Lake Software) which allows Java client code to query the Ingres DART database directly over the internet.

Objective 3 -- DART Transfer

The DART internal disk configuration was upgraded by the addition of eighteen new disks provided by BPA.

An Oracle version 8(i) installation was received from Oracle in October and the implementation was completed. Additional software packages were added including the Oracle Enterprise Manager GUI, the Oracle Developer and a setup configuration to complete the installation.

Migration of the database began manually just after Thanksgiving. An attempted effort to use third party software for the conversion was not successful due to version differences between it and the new 8(i) version.

The major DART data tables were successfully transferred by January 1, 2000.

Judy Cress and Chris Van Holmes attended Oracle training classes for Backup and Recovery, Networking, and PLSQL.

Jutith Cress provided BPA with recommendations concerning a transfer of DART to ORACLE in the first quarterly report of this year.

Problems Encountered during the Year

Hardware disk failures required migration of DART from confocal to gandhi, the Solaris Enterprise 450 server during the year. The decision was based upon lack of backup disk hardware requiring the database to be straddled across our TCP-IP network. The resulting retardation of response time and interruption of other research activities due to network overload revealed that this configuration was not compatible with efficient in-house as well as DART operations. Migrating DART to a single dedicated location with database and all disks mounted on the same server was the most ideal configuration.

A reduced backup schedule was implemented due to a tape drive failure in November and a full backup restored in January when the tape drive hardware was restored.

Hardware Support

Four additional disks were mounted in early October, 1999.

Tape drive and disk replacement as well as several reconfigurations were required during the year due to hardware failures. Aging hardware was phased out and migration to a newer server ultimately provided the answer to most of the associated problems with performance and work interruptions. The tape backup was repaired and a new one purchased to resolve the potential backup deficiencies.

Software Support

Third party conversion software was not supported for the current version of Oracle and had to be abandoned in favor of manually moving the data across to the Oracle Database. The major tables were transferred by the end of December, 1999.

Commercial Software Maintained:

- Ingres (database management)
- Oracle 8 Enterprise Edition
- Developer/2000
- Oracle Enterprise Manager
- Webworks Publisher
- JSVR for Ingres

Public Domain Software Maintained

- Netscape (World Wide Web browser)

Objective 4 -- Regional Info. Management Review

We developed and expanded capabilities in DART to provide summary release and observation data based on a wide range of release criteria per ISAB request. Query output now includes a datafile suitable for importation to SURPH, allowing survival estimates for any chosen release grouping. This capability is being expanded to provide travel time information as well.

A report on DART Query Usage Statistics is also included. (See Appendix A).

We also provided the IRSP with a report entitled “Primary and Secondary Site Level Data Integrity Control Issues,” by Judith Cress. (See Appendix B).

Appendix A: DART Query Usage Statistics for 2000

OCTOBER 1999	Total	Gov	%Gov	%Mil	%US	%IP	%Edu	%Org
Graphic Queries								
River	**6301**		1.46%	0.24%	1.27%	4.14%	1.79%	0.03%
Adult	**5018**		7.59%	0.00%	4.46%	21.02%	3.21%	0.28%
ESU	**4**		25.00%	0.00%	0.00%	0.00%	0.00%	0.00%
PIT Tag	**585**		0.00%	86.15%	0.34%	10.77%	1.20%	0.00%
Smolt	**177**		12.43%	0.00%	4.52%	50.28%	7.91%	0.00%
Climate								
Total Graphic Queries	**12085**		**4.10%**	**4.29%**	**2.60%**	**12.15%**	**2.44%**	**0.13%**
Detail & Composite Reports								
River	**2099**		4.19%	17.44%	0.52%	3.95%	3.38%	4.48%
Adult	**7838**		4.76%	1.89%	2.37%	17.72%	6.74%	0.48%
ESU	**164**		7.93%	51.22%	0.61%	25.00%	3.05%	0.00%
Hatchery	**115**		9.57%	0.00%	1.74%	23.48%	15.65%	0.00%
PIT Tag	**315**		3.81%	0.00%	18.41%	41.59%	10.79%	2.54%
Smolt	**203**		4.43%	59.11%	3.45%	12.81%	1.48%	0.00%
Climate								
CWT								
Total Report Queries	**10734**		**4.71%**	**6.69%**	**2.47%**	**15.81%**	**6.14%**	**1.30%**
Total Queries	**22819**		**4.39%**	**5.42%**	**2.54%**	**13.87%**	**4.18%**	**0.68%**
HTML requests	**10900**		**6.85%**	**0.45%**	**2.99%**	**21.35%**	**6.57%**	**1.06%**

NOVEMBER 1999	Total	Gov	%Gov	%Mil	%US	%IP	%Edu	%Org
Graphic Queries								
River	**5542**		3.72%	1.08%	0.88%	7.25%	5.97%	0.04%
Adult	**2162**		4.95%	0.00%	6.85%	26.36%	9.71%	0.79%
ESU	**67**		10.45%	0.00%	10.45%	65.67%	7.46%	0.00%
PIT Tag	**20**		0.00%	0.00%	0.00%	40.00%	20.00%	0.00%
Smolt	**108**		5.56%	1.85%	1.85%	35.19%	36.11%	0.00%
Climate								
Total Graphic Queries	**7899**		**4.13%**	**0.78%**	**2.61%**	**13.44%**	**7.46%**	**0.24%**
Detail & Composite Reports								
River	**1796**		2.17%	12.08%	0.61%	11.19%	3.01%	1.61%
Adult	**2961**		4.49%	3.01%	5.64%	20.57%	5.81%	0.78%
ESU	**55**		1.82%	0.00%	5.45%	30.91%	32.73%	0.00%
Hatchery	**107**		1.87%	0.00%	0.00%	51.40%	28.97%	0.00%
PIT Tag	**164**		24.39%	0.00%	6.10%	23.17%	18.29%	0.00%
Smolt	**88**		2.27%	1.14%	1.14%	50.00%	10.23%	0.00%
Climate								
CWT								
Total Report Queries	**5171**		**4.20%**	**5.94%**	**3.71%**	**18.64%**	**6.07%**	**1.01%**
Total Queries	**13070**		**4.15%**	**2.82%**	**3.05%**	**15.50%**	**6.91%**	**0.54%**
HTML requests	**7731**		**6.36%**	**0.62%**	**3.69%**	**19.79%**	**12.93%**	**0.93%**

DECEMBER 1999	Total	Gov	%Gov	%Mil	%US	%IP	%Edu	%Org
Graphic Queries								
River	**6853**		4.47%	0.06%	1.33%	8.70%	4.41%	0.07%

Adult	**1297**		2.85%	1.16%	7.94%	22.36%	10.41%	1.23%
ESU	**17**		23.53%	0.00%	47.06%	17.65%	5.88%	0.00%
PIT Tag	**19**		10.53%	0.00%	0.00%	21.05%	63.16%	0.00%
Smolt	**175**		36.00%	0.00%	0.00%	6.86%	35.43%	0.00%
Climate								
Total Graphic Queries	**8361**		**4.93%**	**0.23%**	**2.42%**	**10.82%**	**6.12%**	**0.25%**
Detail & Composite Reports								
River	**2120**		2.74%	23.25%	0.33%	6.27%	2.97%	1.23%
Adult	**1132**		3.27%	18.64%	8.04%	11.84%	6.45%	1.86%
ESU	**30**		3.33%	0.00%	16.67%	50.00%	3.33%	0.00%
Hatchery	**79**		7.59%	0.00%	2.53%	48.10%	20.25%	0.00%
PIT Tag	**101**		23.76%	0.00%	26.73%	17.82%	18.81%	0.99%
Smolt	**180**		11.67%	75.56%	3.33%	3.33%	3.89%	0.00%
Climate								
CWT								
Total Report Queries	**3642**		**4.04%**	**23.06%**	**3.79%**	**9.45%**	**4.91%**	**1.32%**
Total Queries	**12003**		**4.66%**	**7.16%**	**2.83%**	**10.41%**	**5.76%**	**0.57%**
HTML requests	**6189**		**5.74%**	**0.37%**	**4.22%**	**20.31%**	**9.58%**	**1.08%**

JANUARY 2000	**Total**	**Gov**	**%Gov**	**%Mil**	**%US**	**%IP**	**%Edu**	**%Org**
Graphic Queries								
River	**5920**		1.77%	0.27%	1.33%	6.84%	1.79%	0.25%
Adult	**1233**		18.33%	0.24%	4.22%	15.57%	10.54%	0.89%
ESU	**0**		0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
PIT Tag	**15**		0.00%	0.00%	0.00%	86.67%	6.67%	0.00%
Smolt	**225**		7.56%	0.00%	0.00%	50.22%	26.22%	8.89%
Climate								
Total Graphic Queries	**7393**		**4.71%**	**0.26%**	**1.77%**	**9.78%**	**4.00%**	**0.62%**
Detail & Composite Reports								
River	**507**		12.23%	0.20%	3.55%	30.97%	17.36%	5.13%
Adult	**698**		7.45%	0.43%	2.58%	26.93%	9.60%	0.72%
ESU	**57**		1.75%	0.00%	1.75%	68.42%	17.54%	0.00%
Hatchery	**164**		22.56%	0.00%	0.61%	58.54%	4.88%	0.00%
PIT Tag	**107**		30.84%	0.00%	3.74%	39.25%	12.15%	0.00%
Smolt	**134**		4.48%	0.00%	0.75%	82.84%	2.99%	0.00%
Climate	**13**		69.23%	0.00%	0.00%	0.00%	30.77%	0.00%
CWT								
Total Report Queries	**1680**		**11.90%**	**0.24%**	**2.56%**	**37.68%**	**11.55%**	**1.85%**
Total Queries	**9073**		**6.04%**	**0.25%**	**1.92%**	**14.95%**	**5.40%**	**0.85%**
HTML requests	**5120**		**8.50%**	**0.45%**	**2.83%**	**23.50%**	**6.70%**	**1.46%**

FEBRUARY 2000	**Total**	**Gov**	**%Gov**	**%Mil**	**%US**	**%IP**	**%Edu**	**%Org**
Graphic Queries								
River	**6692**		2.44%	0.51%	1.37%	10.16%	2.79%	0.06%
Adult	**1329**		6.92%	2.56%	12.64%	7.67%	2.48%	1.81%
ESU	**23**		4.35%	0.00%	0.00%	34.78%	43.48%	0.00%
PIT Tag	**55**		5.45%	0.00%	1.82%	36.36%	41.82%	5.45%
Smolt	**287**		28.57%	25.44%	2.09%	16.38%	8.36%	2.79%
Climate	**7**		14.29%	0.00%	0.00%	28.57%	14.29%	0.00%
Total Graphic Queries	**8393**		**4.07%**	**1.68%**	**3.18%**	**10.23%**	**3.31%**	**0.46%**

Detail & Composite Reports								
River	**1033**		9.58%	44.43%	1.74%	15.68%	2.61%	1.55%
Adult	**1301**		2.61%	13.91%	1.08%	34.74%	4.69%	0.15%
ESU	**57**		40.35%	0.00%	1.75%	1.75%	24.56%	3.51%
Hatchery	**50**		74.00%	0.00%	0.00%	38.00%	10.00%	0.00%
PIT Tag	**337**		45.10%	0.30%	3.86%	15.43%	20.77%	7.72%
Smolt	**97**		16.49%	11.34%	0.00%	19.59%	1.03%	30.93%
Climate	**30**		33.33%	0.00%	0.00%	33.33%	26.67%	0.00%
CWT	**41**		51.22%	0.00%	12.20%	9.76%	7.32%	0.00%
Total Report Queries	**2946**		**13.31%**	**22.13%**	**1.73%**	**24.41%**	**6.42%**	**2.58%**
Total Queries	**11339**		**6.47%**	**6.99%**	**2.80%**	**13.92%**	**4.12%**	**1.01%**
HTML requests	**7621**		**7.27%**	**1.17%**	**2.70%**	**22.28%**	**8.38%**	**1.27%**

MARCH 2000	**Total**	**Gov**	**%Gov**	**%Mil**	**%US**	**%IP**	**%Edu**	**%Org**
Graphic Queries								
River	**7240**		4.96%	0.17%	1.59%	8.66%	3.37%	0.36%
Adult	**2606**		8.33%	0.50%	8.52%	27.55%	2.88%	0.35%
ESU	**19**		0.00%	0.00%	0.00%	52.63%	5.26%	0.00%
PIT Tag	**29**		3.45%	0.00%	0.00%	44.83%	44.83%	0.00%
Smolt	**207**		16.91%	0.00%	0.00%	69.57%	8.21%	0.00%
Climate	**4**		0.00%	0.00%	0.00%	25.00%	0.00%	0.00%
Total Graphic Queries	**10105**		**6.06%**	**0.25%**	**3.33%**	**14.97%**	**3.46%**	**0.35%**
Detail & Composite Reports								
River	**3515**		1.68%	1.79%	1.79%	4.01%	1.56%	0.28%
Adult	**2696**		6.75%	1.08%	5.16%	23.85%	9.38%	0.74%
ESU	**43**		0.00%	2.33%	2.33%	51.16%	9.30%	0.00%
Hatchery	**64**		31.25%	0.00%	0.00%	25.00%	14.06%	6.25%
PIT Tag	**194**		31.44%	0.00%	5.67%	37.11%	6.70%	2.58%
Smolt	**77**		16.88%	2.60%	6.49%	37.66%	9.09%	15.58%
Climate	**22**		22.73%	0.00%	0.00%	40.91%	22.73%	0.00%
CWT	**50**		40.00%	0.00%	6.00%	26.00%	8.00%	0.00%
Total Report Queries	**6661**		**5.40%**	**1.43%**	**3.33%**	**14.19%**	**5.25%**	**0.77%**
Total Queries	**16766**		**5.80%**	**0.72%**	**3.33%**	**14.66%**	**4.18%**	**0.51%**
HTML requests	**10379**		**6.38%**	**0.44%**	**3.71%**	**29.38%**	**5.63%**	**1.03%**

APRIL 2000	**Total**	**Gov**	**%Gov**	**%Mil**	**%US**	**%IP**	**%Edu**	**%Org**
Graphic Queries								
River	**6384**		3.62%	0.13%	2.05%	5.89%	2.63%	0.39%
Adult	**8348**		7.25%	0.71%	5.28%	22.87%	8.31%	0.16%
ESU	**32**		65.63%	0.00%	0.00%	12.50%	15.63%	0.00%
PIT Tag	**70**		4.29%	51.43%	2.86%	15.71%	0.00%	0.00%
Smolt	**661**		0.15%	0.00%	0.00%	34.49%	60.21%	0.00%
Climate	**28**		21.43%	0.00%	0.00%	17.86%	60.71%	0.00%
Total Graphic Queries	**15523**		**5.59%**	**0.66%**	**3.70%**	**16.32%**	**8.26%**	**0.24%**
Detail & Composite Reports								
River	**2548**		2.59%	6.44%	1.92%	10.71%	6.16%	2.75%
Adult	**10069**		8.60%	0.45%	1.68%	19.62%	7.04%	0.41%
ESU	**247**		30.36%	2.43%	0.00%	53.04%	8.50%	0.00%
Hatchery	**220**		7.73%	0.00%	0.00%	40.45%	1.82%	0.45%
PIT Tag	**652**		16.41%	0.31%	5.83%	50.92%	5.98%	2.15%

Smolt	**172**		22.67%	11.05%	5.81%	29.07%	3.49%	0.00%
Climate	**60**		10.00%	1.67%	0.00%	36.67%	16.67%	0.00%
CWT	**65**		24.62%	0.00%	0.00%	20.00%	21.54%	0.00%
Total Report Queries	**14033**		**8.49%**	**1.69%**	**1.90%**	**20.57%**	**6.84%**	**0.90%**
Total Queries	**29556**		**6.97%**	**1.15%**	**2.84%**	**18.33%**	**7.59%**	**0.55%**
HTML requests	**17893**		**9.04%**	**0.79%**	**2.54%**	**24.59%**	**7.27%**	**0.78%**

MAY 2000	**Total**	**Gov**	**%Gov**	**%Mil**	**%US**	**%IP**	**%Edu**	**%Org**
Graphic Queries								
River	**6932**		4.18%	0.45%	1.73%	21.78%	2.06%	0.16%
Adult	**11011**		4.98%	1.43%	3.83%	21.02%	4.21%	1.77%
ESU	**76**		56.58%	0.00%	0.00%	31.58%	0.00%	0.00%
PIT Tag	**148**		3.38%	6.76%	0.00%	19.59%	7.43%	51.35%
Smolt	**360**		18.06%	0.00%	0.00%	48.89%	18.33%	4.44%
Climate	**33**		0.00%	12.12%	0.00%	3.03%	27.27%	9.09%
Total Graphic Queries	**18560**		**5.12%**	**1.09%**	**2.92%**	**21.85%**	**3.73%**	**1.62%**
Detail & Composite Reports								
River	**2967**		4.68%	12.74%	3.07%	14.96%	6.40%	6.03%
Adult	**12389**		6.43%	1.19%	2.17%	22.56%	6.52%	1.05%
ESU	**290**		32.76%	7.59%	2.07%	44.83%	2.41%	0.34%
Hatchery	**270**		4.44%	0.00%	0.37%	13.33%	9.26%	8.15%
PIT Tag	**1026**		8.58%	3.02%	1.75%	41.81%	5.26%	11.11%
Smolt	**254**		33.07%	5.12%	0.00%	22.05%	7.09%	2.36%
Climate	**61**		0.00%	3.28%	1.64%	27.87%	9.84%	21.31%
CWT	**70**		0.00%	18.57%	2.86%	11.43%	17.14%	7.14%
Total Report Queries	**17327**		**7.01%**	**3.50%**	**2.24%**	**22.59%**	**6.46%**	**2.71%**
Total Queries	**35887**		**6.04%**	**2.26%**	**2.59%**	**22.21%**	**5.05%**	**2.15%**
HTML requests	**23885**		**7.39%**	**1.75%**	**2.12%**	**22.56%**	**5.70%**	**1.61%**

JUNE 2000	**Total**	**Gov**	**%Gov**	**%Mil**	**%US**	**%IP**	**%Edu**	**%Org**
Graphic Queries								
River	**7140**		3.67%	0.63%	2.61%	10.41%	0.94%	0.13%
Adult	**8401**		6.65%	1.19%	5.93%	12.83%	2.61%	4.14%
ESU	**59**		72.88%	0.00%	0.00%	5.08%	0.00%	11.86%
PIT Tag	**71**		0.00%	0.00%	0.00%	12.68%	25.35%	39.44%
Smolt	**172**		26.16%	0.00%	0.00%	29.65%	8.14%	0.00%
Climate	**65**		3.08%	0.00%	3.08%	73.85%	10.77%	3.08%
Total Graphic Queries	**15908**		**5.73%**	**0.91%**	**4.31%**	**12.14%**	**2.04%**	**2.48%**
Detail & Composite Reports								
River	**2885**		3.78%	15.91%	4.30%	7.94%	1.11%	3.50%
Adult	**8341**		8.69%	0.59%	3.86%	25.04%	5.89%	1.09%
ESU	**342**		22.81%	2.34%	0.29%	61.40%	2.63%	0.00%
Hatchery	**108**		17.59%	0.00%	0.93%	46.30%	6.48%	0.00%
PIT Tag	**883**		9.63%	9.63%	1.93%	41.45%	8.72%	4.19%
Smolt	**176**		28.41%	2.84%	0.00%	52.84%	1.70%	4.55%
Climate	**37**		5.41%	13.51%	0.00%	21.62%	16.22%	5.41%
CWT	**532**		0.38%	0.00%	0.00%	4.32%	0.38%	0.19%
Total Report Queries	**13304**		**8.04%**	**4.59%**	**3.50%**	**23.06%**	**4.71%**	**1.80%**
Total Queries	**29212**		**6.78%**	**2.59%**	**3.94%**	**17.12%**	**3.26%**	**2.17%**
HTML requests	**16819**		**8.07%**	**1.88%**	**3.74%**	**26.48%**	**5.89%**	**1.90%**

JULY 2000	Total	Gov	%Gov	%Mil	%US	%IP	%Edu	%Org
Graphic Queries								
River	**6849**		2.12%	0.34%	2.16%	6.41%	0.77%	0.06%
Adult	**7833**		8.16%	0.96%	6.87%	14.20%	2.67%	0.83%
ESU	**89**		80.90%	0.00%	0.00%	15.73%	0.00%	0.00%
PIT Tag	**63**		14.29%	0.00%	0.00%	50.79%	0.00%	28.57%
Smolt	**179**		52.51%	0.56%	2.79%	18.99%	7.26%	0.00%
Climate	**15**		13.33%	0.00%	6.67%	73.33%	6.67%	0.00%
Total Graphic Queries	**15028**		**6.39%**	**0.66%**	**4.60%**	**10.93%**	**1.84%**	**0.58%**
Detail & Composite Reports								
River	**2904**		2.38%	18.25%	2.03%	4.55%	1.07%	3.68%
Adult	**10134**		6.55%	0.85%	3.98%	22.69%	3.34%	1.11%
ESU	**288**		7.29%	13.19%	1.04%	62.50%	1.04%	0.00%
Hatchery	**118**		13.56%	0.00%	0.00%	26.27%	27.12%	3.39%
PIT Tag	**950**		12.42%	17.47%	4.53%	40.11%	5.47%	2.53%
Smolt	**108**		26.85%	11.11%	3.70%	17.59%	11.11%	0.00%
Climate	**51**		0.00%	1.96%	17.65%	13.73%	7.84%	5.88%
CWT	**44**		27.27%	0.00%	0.00%	20.45%	11.36%	4.55%
Total Report Queries	**14597**		**6.36%**	**5.71%**	**3.57%**	**20.95%**	**3.27%**	**1.73%**
Total Queries	**29625**		**6.38%**	**3.15%**	**4.09%**	**15.86%**	**2.54%**	**1.14%**
HTML requests	**18180**		**6.47%**	**1.60%**	**3.46%**	**20.27%**	**8.18%**	**1.46%**

AUGUST 2000	Total	Gov	%Gov	%Mil	%US	%IP	%Edu	%Org
Graphic Queries								
River	**6925**		2.93%	0.78%	2.66%	6.18%	1.60%	1.53%
Adult	**9608**		6.67%	0.62%	6.79%	14.62%	2.39%	2.11%
ESU	**115**		53.04%	0.00%	0.00%	33.04%	4.35%	0.87%
PIT Tag	**44**		6.82%	0.00%	0.00%	54.55%	11.36%	25.00%
Smolt	**201**		60.20%	1.00%	1.49%	30.85%	4.98%	0.00%
Climate	**56**		14.29%	3.57%	0.00%	28.57%	5.36%	25.00%
Total Graphic Queries	**16949**		**6.12%**	**0.70%**	**4.95%**	**11.64%**	**2.15%**	**1.98%**
Detail & Composite Reports								
River	**2841**		6.62%	14.99%	1.97%	5.07%	2.68%	3.27%
Adult	**12261**		7.95%	0.51%	3.27%	20.45%	4.15%	0.55%
ESU	**197**		12.18%	6.09%	0.51%	65.99%	3.05%	1.02%
Hatchery	**27**		3.70%	3.70%	0.00%	18.52%	7.41%	0.00%
PIT Tag	**943**		11.98%	16.65%	4.14%	40.93%	9.54%	0.95%
Smolt	**120**		25.00%	39.17%	1.67%	13.33%	10.00%	0.83%
Climate	**78**		21.79%	0.00%	5.13%	26.92%	8.97%	2.56%
CWT	**88**		5.68%	3.41%	0.00%	18.18%	7.95%	3.41%
Total Report Queries	**16555**		**8.17%**	**4.28%**	**3.04%**	**19.48%**	**4.28%**	**1.08%**
Total Queries	**33504**		**7.13%**	**2.47%**	**4.01%**	**15.51%**	**3.20%**	**1.53%**
HTML requests	**20652**		**8.03%**	**1.23%**	**3.10%**	**21.66%**	**4.72%**	**1.52%**

SEPTEMBER 2000	Total	Gov	%Gov	%Mil	%US	%IP	%Edu	%Org
Graphic Queries								
River	**6693**		2.30%	0.37%	1.46%	3.84%	1.03%	0.85%
Adult	**11998**		8.37%	0.77%	5.93%	10.25%	2.79%	3.64%
ESU	**20**		90.00%	0.00%	0.00%	0.00%	5.00%	0.00%

PIT Tag	**29**	0.00%	3.45%	3.45%	55.17%	13.79%	0.00%
Smolt	**60**	25.00%	5.00%	0.00%	45.00%	11.67%	8.33%
Climate	**33**	42.42%	0.00%	6.06%	9.09%	6.06%	3.03%
Total Graphic Queries	**18833**	**6.40%**	**0.64%**	**4.32%**	**8.14%**	**2.22%**	**2.65%**
Detail & Composite Reports							
River	**2972**	4.48%	16.99%	1.11%	2.09%	2.02%	2.52%
Adult	**15059**	5.43%	0.56%	3.95%	18.74%	5.20%	0.58%
ESU	**74**	24.32%	0.00%	1.35%	40.54%	6.76%	0.00%
Hatchery	**112**	31.25%	0.00%	0.00%	10.71%	5.36%	0.00%
PIT Tag	**822**	11.44%	8.15%	2.55%	37.35%	7.18%	6.20%
Smolt	**70**	58.57%	2.86%	1.43%	4.29%	1.43%	1.43%
Climate	**52**	15.38%	0.00%	11.54%	3.85%	13.46%	0.00%
CWT	**48**	12.50%	0.00%	4.17%	14.58%	0.00%	2.08%
Total Report Queries	**19209**	**6.00%**	**3.43%**	**3.43%**	**16.89%**	**4.79%**	**1.12%**
Total Queries	**38042**	**6.20%**	**2.05%**	**3.87%**	**12.56%**	**3.52%**	**1.88%**
HTML requests	**21696**	**6.12%**	**1.16%**	**3.36%**	**17.27%**	**5.04%**	**1.64%**
Total Graphic Queries	**155037**	**5.46%**	**1.02%**	**3.61%**	**13.03%**	**3.66%**	**1.18%**
Total Report Queries	**125859**	**6.94%**	**4.98%**	**2.95%**	**19.62%**	**5.32%**	**1.49%**
Total Queries 2000	**280896**	**6.12%**	**2.79%**	**3.32%**	**15.98%**	**4.41%**	**1.32%**

Appendix B: Special report by Judith Cress

Primary and Secondary Site Level
Data Integrity Control Issues

Primary data sites represent the resource on which secondary sites and users depend for data integrity. Errant data, frequently the result of data that has escaped controls at the originating site, is passed through inheritance (by publication or electronic transmission), to the secondary site and subsequently onto the user. These errors go largely unnoticed until a point when the data is needed and acquired by the user. It is then the ensuing small (perhaps large) disaster begins.

Unfortunately the user must wrestle with bad data first. Slowed by constraints imposed by the errors, the user must first understand and evaluate the bad data, determine a correction strategy, and if none exists, then bear the contact time involved searching for a corrected data set. The cost in time researching corrections, manual edits or obtaining corrected data files (all considered soft costs), can be far reaching when multiplied by every user affected within a distributed system. Further, if no remedy is found, the correction burden is often passed back to the secondary source which must evaluate, research to determine the correct action, then act. Unless the error was generated at the current site, the secondary site must continue back to the originating source of the error. The broad impact of poor data control looks something like this.

1) Small data errors are embarrassing
    - impacts credibility of everyone passing the data
    - correction required for originating site, subsequent site and user

2) Not-so-large data disasters are costly
    - cause embarrassment to originating site and lead to bad publicity
    - every data user involved must take some defensive action
    - all sites suffers time lost when answering incoming data complaints
    - work predicated on bad data will need reworking at every site level

3) Large data disasters may put the survival of the organization at risk

Bad data does not have an easy remedy. Without integrity controls, obvious data errors live as true and correct values impacting every subsequent transaction employing the data. Similar to a business document such as a title, billing or receipt,

errors must be corrected by the originating party or the integrity cannot be relied upon and subsequent transactions may be voidable. This legal tenant is true with data as well as commercial documents.

Secondary data sites have quite a different set of problems with errant data. Integrity controls omitted by the primary site must be implemented for the sake of their users where incoming data is of poor quality. Creating controls at the secondary level is difficult since the secondary site has no first hand knowledge of the information, but must rely on limited review and ranges for corrections, then remove or log suspected errors. In doing so, large amounts of data must be filtered and reviewed and often smoothed to allow the information to make sense. If periodic updates are provided from the primary site, a major problem arises when the data is corrected by the secondary site, then errors reintroduced because the primary data source has not made corrections - thus the corrected data is overwritten by a new batch of bad data.

How can the primary data site create integrity controls which reduce the costly overhead associated distributing poor quality data? The three following suggestions represent simple beginning to setting controls in place:

1) Raising consciousness of primary site personnel
2) Implementing a phase related control system
3) Create a quality assurance position

Raising Personnel Consciousness

The most effective way to handle data integrity errors is by raising the data integrity consciousness of the primary (and secondary) site personnel. Recognizing the importance of clean data collection, and assuring errors do not pass to other levels without review/correction procedures must become part of the organization's culture. This may require programmed training or hands on change management efforts to involve all personnel in an effort to improve data quality.

In connection with the above, the need for personnel to realize where corrections must be implemented to ensure erroneous data does not creep back into the system after corrections have been made is critical. If the data is corrected in the secondary database, but not on at the original source - the error continues to be distributed despite efforts of the secondary site. Therefore a formal correction strategy must be well understood by data personnel when information processes are distributed across department and organizational levels.

Phase Related Control System

Implementation of a phase-related control system is an effective approach to improving data integrity. Breaking down the data flow into logical segments is the first step, then identifying the best location for integrity controls within each segment.

By identifying the data flow between the various departments such as field collection, preprocessing, database loading, and publication, the logical integrity control placement becomes clear within each department.

The attached flow chart shows the data flow from origination to distribution and examples where controls might be implemented on each level in a phased control system.

Errors can occur at every level of the system, so the focus becomes trapping and correcting before the offending data before it flows to the next level. If data passes to the next level in a condition that does not require a return to a previous level, the control system has succeeded. Since each primary site is different, how and where controls are place requires the analyst eye, and a planned strategy and long-range plan for managing change.

Data Quality Personnel

Data quality improvement takes time and a conscientious effort. To ensure a long term commitment to improved data quality, it is important that someone be responsible for quality assurance, given the authority to oversee policy issues, and be provided with resources and tools to monitor, identify and manage the data as it flows through the system. It is a position too often overlooked or marginalized during budget crunches and personnel shuffling.

The data quality position represents the interface between departments and between agencies regarding data access and transfer matters, and is primarily concerned with the details and public relations issues associated with data distribution. A large data center or warehouse is dependent on the quality of its data, so integrity and quality control must be a high priority.

Data Flow Control

Each phase of data flow control (see attached) must respond to errors incurred at the current level. It is imperative to set controls to trap errors prior to moving onto the next level. This early intervention will prevent most problems as they occur and require the least effort and expense to correct from the overall prospective. It also follows that if an error happens in one department, it cannot be corrected by another department. in other words, the originator of the error owns the error.

The hazards of later intervention can be huge. Once data has passed to the next phase and then an error discovered, a retrofit of some manner must be implemented, either by obtaining a new corrected data set or identify the specifics of the correction and inputting the corrected data. It is important that the ownership of the data be identified at each phase and corrections passed forward from that ownership.

Manual corrections are highly risky and should be avoided. Depending on the circumstances, careful policy judgment is involved and care should be taken. Lawsuits against data shops for unauthorized changes have been successful. Many shops do not make alterations to data, even to obvious errors, but return to the primary site for corrected data sets.

Conclusion

Errors happen for a wide variety of reasons. In order to prevent disasters resulting from bad information, specific steps can be built into a project's processes which reduce passing errors to the subsequent users of the data. It is a matter of earliest possible remedy versus ignoring the problem until a major problem is eminent. Implementing preventative measures require a three pronged approach: Raising the consciousness level of all who handle the data. Implementing a phased control approach to ensure data is validated before it leaves a department in its flow to the next user level. And, setting up a responsible oversite person who can act not only a quality control engineer with full correction authority, but serve as a public interface for policy and public relations matters. In short, quality controls and public relations are the keys to a successful information system.