APPENDIX 4

Communications and Data Processing System

APPENDIX 4a

Data Flow Diagram

Communications and Data – Data Flow Diagram – Level 1

Prepared by:
Larry L. Shannon, WRMA
SB2030 Data Collection and Processing Sub-Team
SB2030 Technical Team

Final Revision: October 25, 1999

History: Discussions began early in the week of August 9, 1999 between the AHA Project Team, State Staff, and Various IBM Technical Teams with the goal of resolving some very valid concerns being raised by State Staff and various IBM Teams. These issues centered on the large numbers of files and the large volumes of data planned to be transferred between the AHA Project Team to the workers in the Counties and returning the collected data to the AHA Project Team. The original techniques designed by the AHA Project Team for the dissemination and collection of data for the Project centered on E-mail facilities with attached data files.

During this week, with help from State Staff, IBM Technical Teams in Sacramento and Boulder, and the AHA Project Team, an alternate solution was devised. This solution significantly reduces the total data load on the State Networks, e.g., the CWS/CMS network facilities, by a) directly transferring only ASCII output data between servers and b) compressing the data format for transmission. This document describes, at a high level, the new data transferring strategies.

Many people have been involved in this solution. This document is a high level explanation of the processes, as currently understood by the AHA Project Team. The AHA Project Team invites all readers to seriously study this document and send corrections and suggestions for later revisions. If you are a member of a Team, please read the processes and flows associated with your Team Process. As everyone knows, the time constraints on this project are significant. Please return comments and clarifications, as soon as possible, to Larry Shannon of Walter R. McDonald & Associates, Inc., lshannon@wrma.com, in Phoenix, 1-623-582-8066, Fax: 1-623-582-6933 or Greg Tooman of American Humane, greg@americanhumane.org, in Denver, 1-303-792-9900, Fax: 1-303-792-5333.

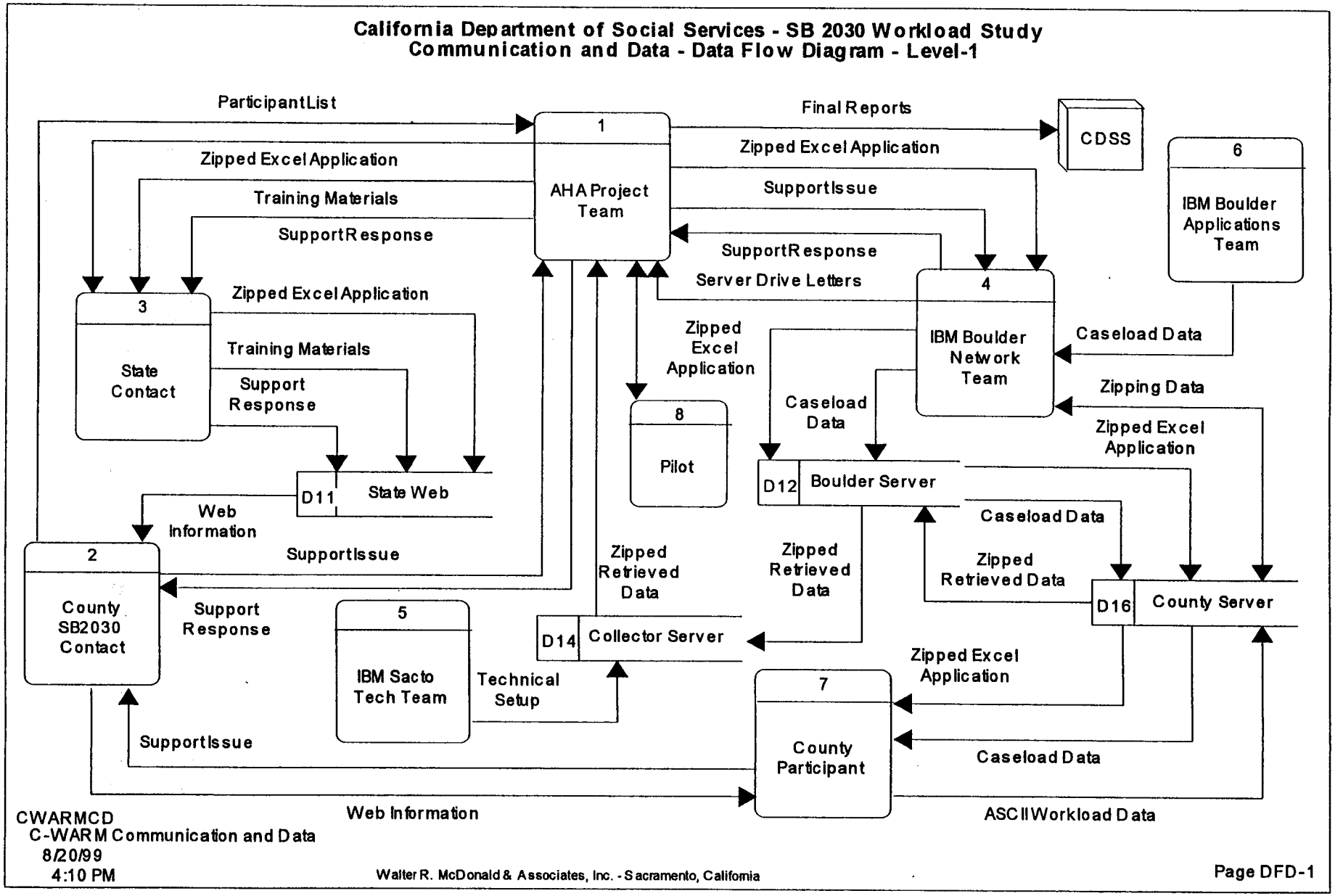
This Document: This technical document shows one selected page from a multi-level Data Flow Diagram (DFD) which focuses on Network Communications, Data Dissemination, and Data Collection for the Workload Study.

The processes in the DFD have been defined around the various Teams of the Project. This will assist the reader in focusing in on the various areas of responsibility, by Team. It is intended that this document will give all Teams on the Project a better understanding of the entire view of the project.

The selected page of the DFD appears on the next page. It may not look good on the screen, but it should print fine. The pages of the document following the diagram give detailed information about the Processes, Data Stores, Data Flows, and External Entities shown on the DFD. Data Flows, both Input and Output, are referenced in the Process sections, but they are described in the Data Flow section. The phrase "*** Not on Diagram ***" indicates a reference exists on a deeper level of the DFD which is not shown on Page 3.

This Version: This version has been updated in retrospect to show more closely how the processes actually worked. The document has also been modified from the CASE Tool output into standard documentation output.

Suggestions for Future Workload Projects: The last section of this document contains Larry's suggestions for improving the project operation in other States.


California Department of Social Services - SB 2030 Workload Study
Communication and Data - Data Flow Diagram - Level-1
Participant List
1
AHA Project Team
Final Reports
CDSS
Zipped Excel Application
Training Materials
Support Response
Support Issue
6
IBM Boulder Applications Team
Server Drive Letters
3
State Contact
4
IBM Boulder Network Team
Caseload Data
Zipping Data
Zipped Excel Application
Support Response
8
Pilot
D11
State Web
D12
Boulder Server
Web Information
2
County SB2030 Contact
Support Issue
Zipped Retrieved Data
Caseload Data
D16
County Server
5
IBM Sacto Tech Team
Technical Setup
D14
Collector Server
7
County Participant
Web Information
ASCII Workload Data
CWARMCD
C-WARM Communication and Data
8/20/99
4:10 PM
Walter R. McDonald & Associates, Inc. - Sacramento, California
Page DFD-1

DATA FLOW DIAGRAM OBJECTS, IN ALPHABETIC ORDER BY NAME
(Refer to the DFD on Page 3)

AHA Project Team Process
Description:
This represents the major processes to be done by the AHA Project Team.
Process #:1
Process Description:
High Level Major Tasks

1. Manage the project.
2. Prepare the training material and train the trainers.
3. Maintain the Workload Flow Document.
4. Coordinate and monitor the Pilot Test.
5. Develop the Excel Application with internal Server Drive Letters.
6. Receive the Participant lists from the Counties.
7. Submit the zipped Excel Application to the IBM Network Team for County distribution and to the State Contact for web publication. See the Zipped Excel Application Section.
8. Develop the data validation and reporting system.
9. Collect, validate, and report on the worker output data files on a daily basis. See the Collection Server Section.
10. Report to the County Contact every day or so concerning the quality of the data being collected and the server status.
11. Prepare a database of the collected data for project analysis.
12. Receive Support Issues.
13. Prepare and distribute Support Responses.

14. Coordinate the Pilot Test.

Location:
C-WARM Communication and Data (0)
Input Flows:
Participant List
Support Response
Support Issue
Zipped Retrieved Data
Server Drive Letters
Output Flows:
Zipped Excel Application
Zipped Excel Application
Support Response
Support Response
Support Issue
Final Reports
Training Materials
Date Last Altered:8/20/99 Date Created:8/12/99

ASCII Workload Data Data Flow

Description:

This represents the Workload Study data collected from all of the Excel Applications for all of the workers. The format is ASCII.

Notes:

This is workload data output from of all worksheets in the Excel Application, saved by the individual worker's Excel Application in ASCII comma separated format onto the County Server. Write permission from the worker's workstation to the County Server is needed.

As the worker "saves" the Personalized Excel Application, all of the data entered to that point in time will be saved in the ccw...w.xls file. When the worker "Closes" the application, custom macros and VB coding will also save an ASCII, Comma Separated Value, file of worker data to the County Server. Example: If the County Server Designated Drive Letter is "M", and the County is Alameda, and the worker's CWS/CMS User ID is ab1234, then the Personalized Excel Application, 01_ab1234.xls, will write this worker's ASCII data output file at: M:\01_ab1234.csv. The worker's Excel Application will be saved at 01_ab1234.xls.

Location:
C-WARM Communication and Data (0)
Source:County Participant (Process)

Dest: County Server (Data Store)
County Participant (7)
Source:*** Not on Diagram ***
Dest: *** Not on Diagram ***
Date Last Altered:8/20/99 Date Created:8/12/99

Boulder Server Data Store
Description:

This represents the CWS/CMS data server in Boulder. It drives the data transfer between the County Servers and the Collection Server.

Data Store #:D12

Notes:

This server is utilized for the following functions:

1. The server is the network source server for the zipped Excel Application distribution to the County destination workstations, at the beginning of each Zone collection period.

2. The server is the network source server for the individual worker caseload files distributed to the County destination servers, at the beginning of each Zone collection period.

3. The server is the network flow-through server for the zipped worker level Excel Application ASCII output data files being swept on a daily basis from the County Servers and transmitted to the Collection Server.

Location:
C-WARM Communication and Data (0)
Input Flows:
Zipped Excel Application
Caseload Data
Zipped Retrieved Data
Output Flows:
Zipped Excel Application
Caseload Data
Zipped Retrieved Data
Date Last Altered:8/20/99 Date Created:8/15/99

Caseload Data Data Flow
Description:

This represents caseload data obtained from CWS/CMS, one worker's caseload per file.

Notes:

Each worker's individual Caseload Data must be available to the worker's workstation when the worker's local Excel Application is launched for the first time. As the Excel Application is launched for the first time, the user will be asked by the Application for the County Number, e.g., 01 = Alameda, their CWS/CMS User ID, and their first and last name.

Custom macros and VB coding within the Application will execute and go to the County Server, e.g., M:\RAWDATA\, to read the caseload file for that given worker. The worker's caseload file will be named @w...w.cld. Example: If the County Server Designated Drive Letter is "M", the County is Alameda, and the worker's CWS/CMS User ID is ab1234, then the Excel Application will look for the worker's caseload file at: M:\RAWDATA\@ab1234.cld.

It will be the responsibility of the IBM Boulder Applications Team to create a program to query the CWS/CMS database to obtain the caseload data for each worker and to create the ASCII, Comma Separated Value, worker caseload files as described and named above.

This Team is also responsible for creating the worker files in the structure described above. However, the creation of these files is by the Zone and only the files for the Counties in a given Zone will be created on the Wednesday night prior to the start of the Zone's collection period.

The IBM Network Team will be responsible for zipping and unzipping, if needed, the above structure, in a manner allow the Excel Application to read the individual files as described above.

The individual worker's workload file is about 200 characters wide and the entire caseload for the State is 185,000.

Both Cases and Referrals will be included in this caseload file.

The Record Layout for the individual worker's caseload, ASCII, Comma Separated Values, is:

1- County ID - 2 characters
2- User ID - 8 characters
3- Staff Person First Name - 20 characters
4- Staff Person Last Name - 25 characters
5- Type of Work - 1 characters (C=Case, R=Referral)
6- Case ID / Referral id - 10 characters
7- Case Name / Referral Name - 35 characters, no internal commas, sorted for ease of finding cases by the worker
8- Count of children in Referral with allegations - 2 characters (null field for a case)
9- Count of children (under 18) in Referral w/o allegations - 2 characters (null field for a case)
10- Case Focus Child First Name - 20 characters (null field for referral)
11- Case Focus Child Middle Name - 20 characters (null field for referral)
12- Case Focus Child Last Name - 25 characters (null field for referral)
13- Client ID - 10 characters

14- Case Focus Child Date of Birth - 10 characters yyyy-mm-dd (null field for referral)
15- Case Focus Child Sex - 1 character (M= Male, F= Female, U = Unknown) (null field for referral)

Location:
C-WARM Communication and Data (0)
Source:IBM Boulder Applications Team (Process)
Dest: IBM Boulder Network Team (Process)
Source:IBM Boulder Network Team (Process)
Dest: Boulder Server (Data Store)
Source:Boulder Server (Data Store)
Dest: County Server (Data Store)
Source:County Server (Data Store)
Dest: County Participant (Process)
IBM Boulder Tech Team (4)
Source:*** Not on Diagram ***
Dest: *** Not on Diagram ***
Source:*** Not on Diagram ***
Dest: *** Not on Diagram ***
IBM Boulder Applications Team (6)
Source:*** Not on Diagram ***
Dest: *** Not on Diagram ***
County Participant (7)
Source:*** Not on Diagram ***
Dest: *** Not on Diagram ***
Date Last Altered:8/20/99 Date Created:8/12/99

CDSS External Entity
Description:
This represents the California Department of Social Services.

Location:
Context Level (CONTEXT)
Input Flows:
Final Reports
C-WARM Communication and Data (0)
Input Flows:
Final Reports
Date Last Altered:8/15/99 Date Created:8/6/99

Collector Server Data Store

Description:

This represents the data Collector Server in Sacramento. It will ultimately receive all of the collected data.

Data Store #:D14

Notes:

This Collector Server resides in Sacramento at the HWDC building in the IBM work area. It will initially be a loaner server, later replaced by a project server. See the Technical Setup Section.

This server will be used for the following functions:

1. Zipped Excel Application data output ASCII, CSV, files are transmitted to this server on a daily basis for the currently operational Zones in the collection schedule. The source of this data is the County Servers.

   These Excel Application output files will be swept from the County Servers and zipped into files named: xccssvvv.zip, where x is an alpha character of "S", "D", or something else, cc is the County ID, ss is the Site ID within the county, and vvv is the Server ID. Example: S3001001.zip = County 30, site 1, server 1.

2. This swept data, by Zone, from the County Servers (all of the xccssvvv.zip files) will be delivered to the Collector Server into path: C:\ddd\, where ddd is the Julian date that the data is dropped into the Collector Server. The files will remain in the zipped format.

3. The WRMA Team will be responsible for removing the data from the Collector Server via the CD writer in the server. The daily data collection will involve only burning the current Julian date directory. Collection will not occur during the weekend and the Monday morning data collection will include the Friday, Saturday and Sunday night server sweeps.

4. After all of the data has been swept, this server will be moved back to the WRMA office in Sacramento. This server did not need to be as big as originally defined. The real power of the server is needed by the WRMA staff in doing initial analysis on the data coming in daily.

Location:

C-WARM Communication and Data (0)

Input Flows:

Zipped Retrieved Data

Technical Setup

Output Flows:

Zipped Retrieved Data

Date Last Altered:8/20/99 Date Created:8/15/99

County Participant Process

Description:

This represents the major processes to be done by the participants providing data to the Study.

Process #:7

Process Description:

The normal method of data entry for the Study is by computer workstation, connected to the CWS/CMS network. If a Free Standing Computer must be used for the data entry, refer to the Use of Free Standing Computers in the County SB2030 Contact Section.

The high level tasks in this process include:

1. Recording Workload Data on Paper Forms

   The County Contact will make a set of paper data collection forms available to all County Participants not having access to a workstation. If possible, these paper forms should be entered into a computer on a daily basis by the County data entry staff. This data entry must be completed within two weeks of the end of the collection period.

2. Frequency of Data Entry into the Excel Application

   Data should be entered into the Excel Application on a daily basis, before the close of the workday, no matter when in the day or night the close of workday is for the worker.

3. Entering Data into the Personalized Excel Application the First Time

   Before starting to enter data for the first time, the County Contact should have made available to the worker's workstation the zipped Master Excel Application file, master.exe. If that file is not present, the worker should contact the County Contact. The worker is responsible for unzipping the master.exe file by double clicking on it from the Explorer.

   The worker should launch Excel using the Excel icon on the computer's desktop. The very first time the worker uses the Excel Application, the worker should open the Master Excel Application on the worker's computer, master.xls.

   As the application is launched, a special Workload Logon screen will appear and request the worker's CWS/CMS User ID or the special unique workload worker ID assigned by the County Contact, if you are not a CWS/CMS user. The Logon screen will also ask for your two digit California County ID. The County ID is necessary because the Excel Application does not yet know for which County you are entering data. The worker clicks OK after the User ID and County ID have been entered.

As soon as this is done, the Master Excel Application will initiate a personalized "save as" of the Excel Application for the worker. The name of the Personalized Excel Application will be cc_w...w.xls, where cc is the County ID and w...w is the worker's CWS/CMS User ID. It is this Personalized Excel Application version that the worker will use from that time on and throughout the Study.

4 Entering Data into the Excel Application After the First Time

The worker should launch Excel using the Excel icon on the computer's desktop. The worker should open the Personalized Excel Application on the worker's computer cc_w...w.xls.

As the application is launched, a special Workload Logon screen will appear and request the worker's CWS/CMS User ID or the special unique workload worker ID assigned by the County Contact, if you are not a CWS/CMS user. The Logon screen will also ask for your County ID, the two digit California County ID.

As soon as this is done, the Personalized Excel Application will be available for the worker to add additional data to the forms.

5 Restarting the Master Application

If the worker desires to discard the current Personalized Excel Application and restart the data entry at Day 1 from the Master Excel Application again, the personalized version of the application cc_w...w.xls, must be deleted from the workstation. After the deletion has occurred, the worker may start at Step 3 again.

6 Saving Workload Data for Submission to the Project Team

The workload data in the Personalized Excel Application file will be saved in two different areas with different formats. When the worker "saves" the Personalized Excel Application file, the Personalized Excel Application file will be saved, as normal, on the worker's local computer cc_w...w.xls. At the end of the data entry session, when the user clicks on "Close", an ASCII file of data will be created and saved on the Designated County Server at M:\cc_w...w.csv, where cc is the County ID and w...w is the worker's CWS/CMS User ID. The "Close" operation may take a minute or two to complete because each of the work sheets in the Personalized Excel Application must be saved separately and then concatenated into the one ASCII file in the server at M:\ccw...w.csv.

7 Entering Data into the Excel Application from Paper Forms

When a data entry person is responsible for the entry of the data from paper forms, the data entry operation is the same as described above, except that care must be taken to use the proper User ID. If an entry person enters 100 workers' data, there will be 100 Personalized versions of the Excel Application on the C:\ drive.

8 Location of the Application and Data on the Workstation

The master Excel Application is pushed down to the workstation and into the desktop folder by the IBM Network team. The actual transfer of the Application occurs when the user logs on and if the Application is not already on the workstation. The personalized version of the Excel Application is also saved to the workstation desktop, as a part of the normal saving of the Excel Application.

When the ASCII worksheet (tab) files are being saved, they are saved to the C:\ drive of the workstation, e.g., Day1.csv. Day2.csv, cover.csv, etc. The total ASCII file, the combination of all of the smaller files is saved to the Server drive at the root. If the server drive is not available, an "access denied" message is presented to the user and the total ASCII file is saved to the workstation C:\ drive.

For future applications of the system, it may not be advisable to use the workstation desktop as the residing place for the Application.

Los Angeles County presented a special twist to this problem of the workstation desktop folder. In LA, the default desktop folder is C:\windows\dcfs\. However, not all of the workstations had this default directory. It was also noted that in many cases, there were both the desktop and the dcsf folders on the workstation. This was the cause of numerous user issues concerning missing Applications, duplicate Applications, etc.

There was also the issue of users being very inventive and moving the Application around to various places and then having multiple copies of the Application present, with varying degrees of data quality, e.g., missing days, etc.

Probably, the Application master, the personalized versions and all of the data should have been kept in one directory, e.g., C:\workload\, with shortcut icons for the master and the personalized version being placed on the desktop.

8 Personal Technical Support for all Workers

The County Contact, or designated assistants, will be the only support contact for the workers in the County. This means that all of the County Participants will need to call the County Contact for all support for programmatic, Excel, computer, and network issues. If the County Contact cannot respond to the issue, the County Contact will call the AHA Support line at 1-877-288-2410. If the AHA Support line cannot resolve the issue, the AHA Support people will contact the IBM Network Support Team.

Location:
C-WARM Communication and Data (0)
- Input Flows:
  - Zipped Excel Application
  - Caseload Data
  - Web Information
- Output Flows:
  - Support Issue
  - ASCII Workload Data

Date Last Altered:8/20/99 Date Created:8/12/99

County SB2030 Contact Process

Description:

This represents the major processes to be done by the County contact person.

Process #:2

Process Description:

The high level tasks in this process include:

GENERAL

1. Prepare the Participant list file for the Project Team.

2. Assist in the training of the County Participants.

3. Assure that the worker caseload files have been received at the County Server. See the Caseload Data Section.

4. Assist the workers in unzipping the Excel Application file at the worker's local workstation.

   The same generic zipped Master Excel Application will have been transmitted to all available workstations by the IBM Network Team at the desktop of the workstation, with the name of master.exe.

   This zipped Excel Application file can be unzipped by double clicking the master.exe file as it is displayed in the Windows Explorer. This unzipping will create the real Master Excel Application file, master.xls. The zipped file will remain on the workstatioin.

5. Assure that Workload Output Data Files are being Saved.

   The County Contact should occasionally check on the County workload server at M:\ to assure that worker output files are being saved properly. These files will be named cc_w...w.csv.

6. State Intranet Support for Workers

   The State Web Site, xxxxxxxxxxxxxx, will be accessible to the County Contact and many of the County Participants. Frequently Asked Questions and Current Issue Resolutions will be available through that web site. The County Contact should visit this site daily to obtain the latest information about the data collection.

7. Personal Technical Support for All Workers

The County Contact, or designated assistants, will be the only support contact for the workers in the County. This means that all of the County Participants will need to call the County Contact for all support for programmatic, Excel, computer, and network issues. If the County Contact cannot respond to the issue, the County Contact will call the AHA Support line at 1-877-288-2410. If the AHA Support line cannot resolve the issue, the AHA Support people will contact the IBM Network Support Team.

PAPER DATA COLLECTION AND ENTRY

Prior to the data collection period for a County, the Project Team will mail a master set of Data Collection Paper Forms to the County Contact. The County Contact should make copies of the forms and then distribute the data forms to all workers needing paper forms.

1. The County Contact will be responsible for providing the paper data collection forms to the workers who do not have a computer.

2. The County Contact will be responsible for seeing that the paper forms are entered into a computer as soon as possible and within two weeks of the close of the data collection period for that County.

3. Caseload data files may or may not be available for workers submitting data by paper. If the file is available, at M:\RAWDATA\@w...w/cld, the Excel Application will access it using the worker User ID at the time of data entry into the application.

USE OF FREE STANDING COMPUTERS

The normal mode of entering data for the Study will be on computers that are connected to the CWS/CMS network. If a Free Standing Computer (FSC) is used for entering the data, the following steps must be taken. The County Contact is responsible for monitoring these steps.

1. The zipped Master Excel Application must be loaded on the FSC. The zipped file can be copied from a networked computer or from the State Web. The zipped form of this file will easily fit onto a 1.4mb diskette.

2. The master.exe file must be unzipped by double clicking on it from the Explorer.

3. The caseload file, if available, should be loaded from the County workload server, M:\RAWDATA\@w...w.cld onto the worker's FSC at C:\rawdata\@w...w.cld. This caseload file should be available for all staff with a CWS/CMS User ID and with primary assignment to some cases. This caseload file will have the CWS/CMS User ID in the file name as shown above (w...w). This file will easily fit onto a 1.4mb diskette.

4. The worker will be asked by the Excel Application if the data entry is being done on a Free Standing Computer. If the answer is yes, then the Excel Application will try to read the caseload file from the local drive at C:\rawdata\@w...w.cld and also write the output data files to the local drive at C:\cc_w...w.csv.

5. The County Contact is responsible for assuring the local data output files at C:\cc_w...w.csv are copied onto the County Workload Server at M:\cc_w...w.csv on a daily basis.

Location:
C-WARM Communication and Data (0)
Input Flows:
Zipped Excel Application
Support Response
Support Response
Support Issue
Training Materials
Web Information
Output Flows:
Participant List
Support Issue
Web Information
Date Last Altered:8/20/99 Date Created:8/12/99

County Server Data Store
Description:
This represents the data servers at the County Site level. They hold caseload data and the collected ASCII output data.

Data Store #:D16
Notes:

1. Drive "M", with only a few exceptions, will be the Designated Drive for most of the Counties. Drive M: will be used in the discussion below. The drive letter is internal to the master Excel Application.

2. There will be a share name on the County Servers of "workload", equivalent to the M:\ drive and the drive will have Read and Write permissions.

3. The caseload worker files sent to the County Servers will be located at M:\RAWDATA\@w...w.cld. They will be unzipped by the IBM Network Team and ready to read into Excel.

4. The Excel Application will be delivered directly to the worker's workstation into file master.exe, BUT will not be unzipped by the IBM Network Team. The User will have to execute this self-extract file, master.exe, to get to the generic Excel Application.

5. The Excel Application data output files, ASCII, CSV, will be saved to M:*.csv by the Excel Application, at Excel Close time.

6. The Excel Application output files will be swept from the County Servers and zipped into files named: xccssvvv.zip, where x is an alpha character of "S", "D", or something else, cc is the County ID, ss is the Site ID within the county, and vvv is the Server ID. Example: S3001001.zip = County 30, site 1, server 1.

7. The swept data from the County Servers (all of the xccssvvv.zip files) will be delivered to the Collector Server into path: C:\ddd\, where ddd is the Julian date of the zipped file on the County Server. The WRMA Team will be responsible for unzipping and processing these xccssvvv.zip files.

For further information, refer to the Zipped Retrieved Data and the Collector Server Sections.

Location:
C-WARM Communication and Data (0)
Input Flows:
Zipped Excel Application
Zipping Data
Caseload Data
ASCII Workload Data
Output Flows:
Zipped Excel Application
Caseload Data
Zipped Retrieved Data
Date Last Altered:8/20/99 Date Created:8/15/99

Final Reports Data Flow
Description:
This represents the set of Final Reports to CDSS.
Location:
Context Level (CONTEXT)
Source:C-WARM Communication and Data (Process)
Dest: CDSS (External Entity)
C-WARM Communication and Data (0)
Source:AHA Project Team (Process)
Dest: CDSS (External Entity)
Project Team (1)
Source:Analyze Data (Process)
Dest: *** Not on Diagram ***
Date Last Altered:8/15/99 Date Created:8/5/99

IBM Boulder Applications Team Process

Description:

This represents the major processes to be done by the IBM Applications Team in Boulder.

Process #:6

Process Description:

The major tasks for this process include:

1. Prepare the programming for the generation of the Worker Caseload files. See the Caseload Data description for the record layout of this file.

2. Periodically, transmit caseload file sets to the IBM Network team for distribution to the County Servers. These sets are transmitted by Zone. A Zone transfer would be for all Counties within the given Zone. The individual worker caseload files are named @w...w.cld, where w...w is the CWSCMS User ID.

3. Participate in the Pilot testing of the data transmissions.

Location:

C-WARM Communication and Data (0)

Output Flows:

Caseload Data

Date Last Altered:8/20/99 Date Created:8/12/99

IBM Boulder Network Team Process

Description:

This represents the major processes to be done by the IBM Network Team in Boulder.

Process #:4

Process Description:

The major tasks for this process include:

CASELOAD DISTRIBUTION

1. Receive the caseload files, by Zone, from the IBM Applications Team. The files are named @w...w.cld. These files are received just before the start of the Zone's collection period.

2. Transmit the caseload files to the County Servers, by the Zone. If the files are zipped for transmission by the Network Team, the files will be unzipped by the Network Team at the County Server. The placement of the files on the County Server will be at

M:\RAWDATA*.cld. These file must be readable from the server by the Excel Application. These files are transmitted just before the start of the Zone's collection period.

3. Participant in the Pilot

DATA COLLECTION SWEEP

4. Zip the data output files on the County Server located at M:\ccw...w.csv into one zipped file named, xccssvvv.zip, where x is an alpha character of "S", "D", or something else, cc is the County ID, ss is the Site ID within the county, and vvv is the Server ID. Example: S3001001.zip = County 30, site 1, server 1.

5. This swept data from the County Servers (all of the xccssvvv.zip files) will be delivered to the Collector Server into path: C:\ddd\, where ddd is the Julian date of the zipped file on the County Server. The files will remain in the zipped format on the Collector Server.

6. Participant in the Pilot testing of the data transmission.

Location:
C-WARM Communication and Data (0)
Input Flows:
Zipped Excel Application
Support Issue
Caseload Data
Output Flows:
Zipped Excel Application
Support Response
Zipping Data
Caseload Data
Server Drive Letters
Date Last Altered:8/20/99 Date Created:8/12/99

IBM Sacto Tech Team Process
Description:
This represents the major processes to be done by the IBM Tech Team in Sacramento.
Process #:5
Process Description:
The major tasks for this process include:

1. Assist in the total coordination of the IBM Teams.

2. Set up the "loaner" server on the network in the IBM office near where Lee Stolmack works. This includes assigning an IP address to the computer and making it work with respect to receiving data from the IBM Network Team.

3. Set up the "production" server on the network in the IBM office near where Lee Stolmack works. This includes assigning an IP address to the computer and making it work with respect to receiving data from the IBM Network Team.

4. Retire the "loaner" server when finished using it.

5. Assist the Collection Team as needed.

6. Participate in the Pilot testing of the data transmission.

Location:
C-WARM Communication and Data (0)
Output Flows:
Technical Setup
Date Last Altered:8/20/99 Date Created:8/12/99

Participant List Data Flow
Description:
This represents a file of various data elements concerning each of the participants in the Study, County employees and non-employees.
Notes:
This is the Participant list file that will be generated by the County Contact prior to the start of the data collection. It is only transmitted from the County Contact to the AHA Project Team. This file will be used for inventory control of data files submitted and data files not submitted by the workers in the County. The format of the file is ASCII, Comma Separated Value, and named ccPartic.csv. The fields in the file are:

Field #1: County Code, 2 characters, numeric, required. This is the California County number.

Field #2: County Staff ID, 20 characters, alphanumeric, required. This is the County's normal identifier for the staff who are County employees. All participants will have this field. For persons who are not County employees, the County Contact must assign a unique County Staff ID.

Field #3: CWS/CMS User ID, 15 characters, alphanumeric. If a participant is a registered user on CWS/CMS, this field is that CWS/CMS User ID.

Field #4: Participant First Name, 20 characters max, alphanumeric, required.

Field #5: Participant Last Name, 25 characters max, alphanumeric, required.

Field #6: Worker Phone Number, 12 characters, required, ###-###-####.

Field #7: E-mail address, 30 characters max.

Field #8: Worker Position Type, 30 characters, alphanumeric, required.

This completed file should be e-mailed to or greg@americanhumane.org or lshannon@wrma.com.

In retrospect, this did not happen. About half of the Counties produced a list of some type, Excel, etc. There should have been more early planning of what was to happen to the County data once it came to the Team. This is a very complex operation if it is to be integrated into the entire data processing operations.

Location:
C-WARM Communication and Data (0)
Source:County SB2030 Contact (Process)
Dest: AHA Project Team (Process)
Project Team (1)
Source:SB2030 Contact (External Entity)
Dest: Prepare Participant Database (Process)
County SB2030 Contact (2)
Source:*** Not on Diagram ***
Dest: *** Not on Diagram ***
Date Last Altered:8/20/99 Date Created:8/12/99

Pilot Process
Description:
This represents the major processes to be done to accomplish a short Pilot Test of the data collection procedures for the Study.
Process #:8
Process Description:
Purpose:
To test the data transmission processes and the data flow of the files across the networks.

Participating Counties:
Merced and San Joaquin, simultaneously.

Numbers of Participants:

Each of the two Pilot Counties should have more than one County Participant, which will result in a completed worksheet each Participant is the Pilot. The Participant may actually be the worker or someone else just doing the Pilot testing.

Data Quality:

The data may be fictitious, but should be formatted correctly.

Test Period:

August 23 and 24, or soon thereafter. The exact schedule is yet to be determined. Each County will enter only two days of workload information into the Excel Application. This may be done in one day.

Minimum Request of the County:

Completion of two days of data for each of two separate workers, each having used a separate copy of the Excel Application.

General Tasks in the Pilot Test:

1. IBM Network Team supplies the Designated Server Drive Letters, for all Counties, to the Project Team for inclusion in the Excel Application.

2. Project Team contacts the two Counties and provides mini-training.

3. Project Team completes the Excel Application, zips the Excel Application into self-extracting format, and e-mails the zipped application file to the IBM Network Team, Walter Arnhols.

4. Project Team will identify a group of CWS/CMS User IDs and accompanying caseload lists, which may be used for the Pilot.

5. IBM Applications Team prepares the caseload files for the predetermined User IDs (or the entire set of workers) in both Pilot Counties and transmits the files to the IBM Network Team, Walter Arnhols.

6. IBM Network Team will transmit the zipped Excel Application only to the predetermined User ID workstations in each County.

7. IBM Network Team will transmit only the caseload files for the predetermined User IDs to the proper County Server.

8. Each Pilot Participant unzip the Excel Application which has been transmitted to their own workstation. (This may be performed by someone other than the identified worker.)

9. Each worker (or other designee) will launch the generic Excel Application and enter data for two days, "Saving", "Closing", and "Exiting" the application for each "day".

10. Each night of the IBM Network Team will zip all of the output data files from the Pilot Counties' servers and transmit the zipped file to the Collector Server in Sacramento.

11. Daily the Project Data Team will copy the data from the Collector Server and make the data available to the AHA Team and the WRMA Team.

12. The AHA and WRMA Teams will validate the data and produce reports on the condition of the data collected.

13. The Pilot processes will be evaluated and enhancements made starting on August 25.

This Pilot effort was conducted in a fairly loose manner. It was originally thought that the Pilot would test all aspects of the processes for the data entry and collection. It actually focused only on the testing of the data transfer of the caseload data to the Counties, the master application to the Counties, and the sweeping of the data from the County servers. This testing of the data transferring was very helpful to the project. It is not clear how helpful the testing of the actual data entry was for the Counties.

Location:
C-WARM Communication and Data (0)
Date Last Altered:8/20/99 Date Created:8/20/99

Server Drive Letters Data Flow

Description:

This represents a list of County Server Drive Designator Letters, pointing to the County/Site Servers to be used to transmit data for the Study.

Notes:

This is information that must be generated by the IBM Boulder Distribution Team and submitted to the Project Team prior to the final version of the Excel Application being distributed. This information is the set of all Designated Server Drive Letters on the County/Site servers which will be used to download and upload data for the Study. The default for the drive letters will be "drive M:\". The Letters will tell the Excel Application the location from which to read the worker's Caseload data for the County (M:\RAWDATA\@w...w.cld) and the location where the Excel Application will save the worker's personalized ASCII output data as M:\cc_w...w.cvs. It is also the location of the nightly data sweep for uploading the collected data to the AHA Project Team.

In retrospect, this list seemed to be an ongoing process and the cause for numerous "crises" for the Project. The entire list could not be completed before the data began to flow. Sometimes last minute changes were needed in the letters, due to bad information. Having the list internal to the Excel Application seemed very restrictive because a change in the letters required a change in the Application. It was felt that the letter choice should not be left to the user. This was a good decision. Possibly an accompanying table of drive letters would be the best way to transmit the drive letters to the field and just deliver the letters file with the Application.

Location:
C-WARM Communication and Data (0)

Source:IBM Boulder Network Team (Process)
Dest: AHA Project Team (Process)
Date Last Altered:8/20/99 Date Created:8/12/99

State Contact Process
Description:
This represents the major processes to be done by the State Contact person.
Process #:3
Process Description:
The major tasks for this process include:

1. Support the project from the State data environment.

2. Assure that certain items are published on the State web, e.g., Frequently Asked Questions, the zipped Excel Application for limited distribution, supplementary training materials, etc.

3. Assist with the coordination of the contractor, the State, and the IBM Teams.

4. Support the data collection team by providing access to the Collector Server as needed.

Location:
C-WARM Communication and Data (0)
Input Flows:
Zipped Excel Application
Support Response
Training Materials
Output Flows:
Zipped Excel Application
Support Response
Training Materials
Date Last Altered:8/20/99 Date Created:8/12/99

State Web Data Store
Description:
This represents a web site where Counties can access information from the State.
Data Store #:D11
Location:
C-WARM Communication and Data (0)

Input Flows:
Zipped Excel Application
Support Response
Training Materials
Output Flows:
Web Information
Date Last Altered:8/15/99 Date Created:8/15/99

Support Issue Data Flow

Description:

This represents a total set of Support Issues, raised by anyone at the County level.

Notes:

This represents any support issue that arises: Excel Application, computer, network, download, etc.

The County Contact, or designated assistants, will be the only support contact for the workers in the County. This means that all of the County Participants will need to call the County Contact for all support for programmatic, Excel, computer, and network issues. If the County Contact cannot respond to the issue, the County Contact will call the AHA Support line at 1-877-288-2410. If the AHA Support line cannot resolve the issue, the AHA Support people will contact the IBM Network Support Team.

Location:
Context Level (CONTEXT)
Source:County Participants (External Entity)
Dest: C-WARM Communication and Data (Process)
C-WARM Communication and Data (0)
Source:County Participant (Process)
Dest: County SB2030 Contact (Process)
Source:AHA Project Team (Process)
Dest: IBM Boulder Network Team (Process)
Source:County SB2030 Contact (Process)
Dest: AHA Project Team (Process)
Project Team (1)
Source:*** Not on Diagram ***
Dest: Project Help Line (Process)
County SB2030 Contact (2)
Source:County Participants (External Entity)
Dest: County Help Line (Process)
IBM Boulder Tech Team (4)
Source:*** Not on Diagram ***

Dest: *** Not on Diagram ***
County Participant (7)
Source:*** Not on Diagram ***
Dest: *** Not on Diagram ***
Date Last Altered:8/20/99 Date Created:8/5/99

Support Response Data Flow

Description:

This represents the solution to Support Issues that have been raised. Some responses will be general and some will be specific.

Notes:

The County Contact, or designated assistants, will be the only support contact for the workers in the County. This means that all of the County Participants will need to call the County Contact for all support for programmatic, Excel, computer, and network issues. If the County Contact cannot respond to the issue, the County Contact will call the AHA Support line at 1-877-282-2410. If the AHA Support line cannot resolve the issue, the AHA Support people will contact the IBM Network Support Team.

Location:
Context Level (CONTEXT)
Source:C-WARM Communication and Data (Process)
Dest: County Participants (External Entity)
Project Team (1)
Source:Project Help Line (Process)
Dest: *** Not on Diagram ***
C-WARM Communication and Data (0)
Source:AHA Project Team (Process)
Dest: State Contact (Process)
Source:State Contact (Process)
Dest: State Web (Data Store)
Source:IBM Boulder Network Team (Process)
Dest: County SB2030 Contact (Process)
Project Team (1)
Source:Project Help Line (Process)
Dest: *** Not on Diagram ***
State Contact (3)
Source:*** Not on Diagram ***
Dest: *** Not on Diagram ***
Source:*** Not on Diagram ***
Dest: *** Not on Diagram ***
IBM Boulder Tech Team (4)

Source:*** Not on Diagram ***
Dest: *** Not on Diagram ***
County Participant (7)
Source:*** Not on Diagram ***
Dest: *** Not on Diagram ***
County SB2030 Contact (2)
Source:County Help Line (Process)
Dest: County Participants (External Entity)
Source:Project Help Team (External Entity)
Dest: County Help Line (Process)
C-WARM Communication and Data (0)
Source:AHA Project Team (Process)
Dest: County SB2030 Contact (Process)
Data Flow --> Web Information
Date Last Altered:8/20/99 Date Created:8/5/99

Technical Setup Data Flow
Description:

This represents the tasks that are needed to set up the Collector Server and make it operational.

Notes:

The IBM Sacramento (Sacto) Tech Team will install, setup, and connect the Collector Server to the State network. This server will be provided by the AHA Project Team. This installation will allow the collection of the Study data, transferred nightly to the collector from the County Servers. The server will have its own IP address, to be determined by the IBM Sacramento Tech Team.

This setup process will be for both the "loaner" and the Production" Collector Server.

Location:
C-WARM Communication and Data (0)
Source:IBM Sacto Tech Team (Process)
Dest: Collector Server (Data Store)
IBM Sacto Tech Team (5)
Source:Setup Collection Server (Process)
Dest: *** Not on Diagram ***
Date Last Altered:8/20/99 Date Created:8/12/99

Training Materials Data Flow
Description:

This represents a general set of training material, some delivered in hardcopy and some via the State web site.

Notes:

These are supplementary training materials. Most of the training material will be provided directly to the Counties via the County Contact. However, many of the materials will also be made available for occasional downloading from a State Web.

Location:
C-WARM Communication and Data (0)
Source:AHA Project Team (Process)
Dest: State Contact (Process)
Source:State Contact (Process)
Dest: State Web (Data Store)
Project Team (1)
Source:Prepare Training Materials (Process)
Dest: State Contact Person (External Entity)
State Contact (3)
Source:*** Not on Diagram ***
Dest: *** Not on Diagram ***
Source:*** Not on Diagram ***
Dest: *** Not on Diagram ***
Date Last Altered:8/20/99 Date Created:8/12/99

Web Information Data Flow
Description:
This represents any information that can be obtained by County level people from a State web site.
Composition:
Zipped Excel Application +
Training Material +
Support Response
Notes:
Web information is the combination of the Zipped Excel Application, the supplementary Training Material, and the Support Response information.

Location:
C-WARM Communication and Data (0)
Source:State Web (Data Store)
Dest: County SB2030 Contact (Process)
Source:County SB2030 Contact (Process)
Dest: County Participant (Process)
County SB2030 Contact (2)

Source:*** Not on Diagram ***
Dest: *** Not on Diagram ***
Date Last Altered:8/20/99 Date Created:8/12/99

Zipped Excel Application Data Flow

Description:

This represents the generic data collection instrument, an Excel application, in self-extracting format.

Notes:

This is the Master Generic Excel Application, in self-extracting format, named master.exe. The compressed size is expected to be between 700kb and 1mb. This file will be created by the AHA Project Team and delivered to the IBM Boulder Network Team via E-mail. It will be transferred to the workstations across the network by the Network Team. This zipped version will be unzipped at the workstation by the worker.

Location:

C-WARM Communication and Data (0)
Source:AHA Project Team (Process)
Dest: State Contact (Process)
Source:State Contact (Process)
Dest: State Web (Data Store)
Source:AHA Project Team (Process)
Dest: IBM Boulder Network Team (Process)
Source:IBM Boulder Network Team (Process)
Dest: Boulder Server (Data Store)
Source:Boulder Server (Data Store)
Dest: County Server (Data Store)
Source:County Server (Data Store)
Dest: County Participant (Process)

Project Team (1)
Source:Prepare Excel Application (Process)
Dest: Self_Extract Excel Application (Data Store)

State Contact (3)
Source:*** Not on Diagram ***
Dest: *** Not on Diagram ***
Source:*** Not on Diagram ***
Dest: *** Not on Diagram ***

IBM Boulder Tech Team (4)
Source:*** Not on Diagram ***
Dest: *** Not on Diagram ***
Source:*** Not on Diagram ***

Dest: *** Not on Diagram ***
County Participant (7)
Source:*** Not on Diagram ***
Dest: *** Not on Diagram ***
Source:*** Not on Diagram ***
Dest: *** Not on Diagram ***
County SB2030 Contact (2)
Source:*** Not on Diagram ***
Dest: *** Not on Diagram ***
Data Flow --> Web Information
Date Last Altered:8/20/99 Date Created:8/12/99

Zipped Retrieved Data Data Flow

Description:

This represents the set of worker data saved from the Excel Application and zipped for transmission to the Project Team.

Notes:

For the current Zones that are collecting data, on a nightly basis, the IBM Boulder Network Team will be responsible for the data compression and data sweep of the *.csv data files from all County Servers for all County Sites for all Counties for all Zones being collected that night. On each server the M:\cc_w...w.csv files will be compressed into one zipped file, named xccssvvv.zip, where x is an alpha character of "S", "D", or something else, cc is the County ID, ss is the Site ID within the County, and vvv is the Server ID within the Site.

Example:
Two zones are being collected tonight. The first Zone has two Counties, (03 and 05) each with one Site and one Server per Site. The zipped files transmitted to the Collector Server from the first Zone are:

D0301001.zip
D0501001.zip

The second Zone being collected contains 3 Counties (ID = 23, 24, 25). Counties 23 and 24 have only one Site with only one Server per Site. County 25 has three Sites with only one Server per Site. The zipped files transmitted to the Collector Server from the second Zone are:

D2301001.zip
D2401001.zip
D2501001.zip
D2502001.zip
D2503001.zip

Location:

C-WARM Communication and Data (0)
Source:County Server (Data Store)
Dest: Boulder Server (Data Store)
Source:Boulder Server (Data Store)
Dest: Collector Server (Data Store)
Source:Collector Server (Data Store)
Dest: AHA Project Team (Process)
Project Team (1)
Source:*** Not on Diagram ***
Dest: Process Data (Process)
Date Last Altered:8/20/99 Date Created:8/12/99

Zipping Data Data Flow

Description:

This represents the compressing or zipping of data by the IBM Network Team in Boulder.

Notes:

This data flow represents the zipping of Caseworker Caseload data files, if needed, into one file per County a few days before the start of the data collection for the County. The zipping will be done by the IBM Boulder Network Team. After transmission to the County Server, the zipped file will be unzipped by the Network Team.

If a County is operating with more than one Site or Server, the same County caseload files must be transmitted to all servers in the County.

The Excel Application at the worker's workstation must be able to read the caseload file into the worker's Excel Application.

Location:
C-WARM Communication and Data (0)
Source:IBM Boulder Network Team (Process)
Dest: County Server (Data Store)
IBM Boulder Tech Team (4)
Source:*** Not on Diagram ***
Dest: *** Not on Diagram ***
Source:*** Not on Diagram ***
Dest: *** Not on Diagram ***
Date Last Altered:8/20/99 Date Created:8/12/99

SUGGESTIONS FOR FUTURE WORKLOAD PROJECTS

This project has been successfully accomplished in an extremely short period of time (mandated by the California State Legislature), under some extreme conditions by a committed Project Team of very good people. Even with all of the "feelings" from the Counties and through the stress of the schedule on the Project Team, the project has been "pulled off" in a superb manner. All who participated in some manner should be congratulated and feel good about the job that has been done.

The State Contact

It is essential that the State have a programmatic and technical (maybe two people) be assigned to the Project. In the case of California, Lee Stolmack was an essential part of the success of the project. She was able to make things happen from the State and County side and also integrate the IBM teams into action.

The County Contacts

I would suggest that there be two County Contacts (CC): one person with the full responsibility and one computer expert with the technical responsibilities. The CC should be committed to the task and not persons that have vacations or "leave of absence" in the middle of the project. In many Counties, the CC was just a figurehead. Fortunately there were usually people in the California Counties who would step up to the tasks and make them happen, whether they were assigned or not. Many times the Project team was never informed who the real (non-figurehead) CC was. All CC should be readily contractible and have e-mail addresses, even if it is with Yahoo.com.

IBM Teams

For the most part, the IBM teams were very helpful and responsive to our needs. It was a very difficult time for them with the concurrent roll out of the CWSCMS Version 4.0. Occasionally, it seemed that some IBM Team members did not follow exactly what we thought we had told them. Maybe this was our communication problem also, e.g., data sweeping, etc. The ability to hold ad hoc and regular open conference calls was very helpful.

The Project Timing

The project time was too short. However, sever County people commented that if much more time have been available, there would still have been a mad rush to completed it at the end. For the technical side of the Project, additional time was definitely needed. Most of the technical functions were designed just days before the production need for the function, e.g., the data processing system, the data reduction and data analysis. There was no total plan for all of the functional aspects of the project.

There is probably no good time to conduct a project of this type. It would be helpful to first review State and County activities to try to avoid vacations, software migrations, etc. The August, school, software roll out, etc. were all interesting hurdles to jump over.

County Training

The County training seemed to go well. It would have been better if all of the training team would have been uniformly trained after the software was finished. This would insure that all of the training of the County people would have been more uniform.

The Data Collection Instrument and Database

The Excel Application was the only way to accomplish the data collection at the County level in California in the time frame we had. However, there was much confusion about the application, some differences in how it was to be used, and issues of error recovery.

In the future, with proper design time, a small data collection application, executable in single user or in a LAN environment, would be better. The application would be written in a standard programming language, e.g., Visual Basic, and linked to a standard database engine, e.g., Access or SQL Server. Time would be given to make the application completely "bullet proof". The database would be designed so that multiple participants could use the application from any workstation without multiple copies of the application being present.

The table of County Server drive letters was another nightmare. In the future, either the servers need to be identified completely and permanently prior to the project's start, or the drive letter table needs to be outside of the application. The need to transmit new versions of the software application just because the drive letters were not correct was a real source of stress for the team and the participants.

The application would also have an "uninstall" feature to not only remove the application from the workstation, but to remove all data file traces. The data transfer to the Project Team would be done in much smaller data files than with the Excel data dumps, possibly allowing the e4-mail data transfer technique.

Workstation Directories, Data Paths and Icons

The mixture of the workstation directories, data paths and the location of icons turned out to be a very difficult problem with much confusion on the part of the Team and the participants. In the future, all of the local software and data files should be installed into a "workload" directory on the users C:\ drive. An application icon should be placed onto the desktop as a "shortcut". There should be no "master, vanilla flavored" application. Whether the icon were installed automatically or manually is not important. Even the manual creation of the shortcut would have been simpler than the nightmare we faced with the "desktop" scenario.

This new solution would also prevent the "multiple application copies" syndrome and most of the lost data issues we faced in California.

Data Flow

It was initially designed that the data from the application would be transmitted to the Project team via the e-mail system. This was a good idea, but assumed that all participants had e-mail and the network lines could sustain the data load. In the future, smaller data files must be transmitted. The data transfer technique used, the "data sweep", turned out to be a good solution. The basis problem here was simply getting all of the servers to be swept on a daily basis, all through the project.

The Collection Server

The Collection Server concept worked well. However, it would have been better if the collection server could have been in the WRMA office. This would have eliminated the need for the daily data collection by WRMA by staff at the State office.

The size of the Collection Server could have been much smaller. The power of the Collection Server (P-III, 550mhz, 256RAM, 6gb, 18gb, 18gb), installed at the State office, would have been more effectively used for the data processing at WRMA. A much smaller server would have worked for collecting the data off the network.

The Data Processing Function

DESIGN: Although the data processing worked well for the project, those functions were designed and implemented in a very fast manner. A comprehensive design was never completed due to the time element, implementation was done along the way, and continued well after the data started to flow. The same is true for the data reduction processes. In the future there needs to be a more thorough design and implementation of the entire data processing from start throughout analysis.

DUPLICATIVE PROCESSING: One of the biggest issues in the data processing was the duplicative processing of worker information that was done for every data sweep. This was due partially to our serious concern that the data transmission would not hold up and we wanted to retain the accumulation of data from each of the daily sweeps. A partial solution to this would be to remove the data from the servers on a nightly basis after the sweep. Additionally, the initiation of a true database engine into the process might also allow incremental data transmissions and processing.

EASE OF DATA REDUCTION: The processes used to produce daily information and feedback to the users concerning the quality of data was complicated enormously by the ASCII format and the duplicative nature of the collected data sets. In the future, the data should be collected into a database at the workstation, incrementally transmitted and immediately validated and inserted into a single master database. This would keep duplicative data processing and storage to a minimum and make data feedback reporting much simpler.

QUOTE MARK PROCESSING: Using Excel for the data collection, in connection with the comma separated value format, and some manual user data entry (not using the table driven data input features in the Application) created may opportunities for data shifts in the record layout. Most of these were recognized and corrected, but still some slipped through.

VALIDATION: True validation of the data set was at a minimum due to the time constraints. Increasing the validation at the data entry level and better design on the processing tools would help in the future to standardize this validation process.

Data Quality Assurance

Knowing When to Shut Servers Off

The Data Analysis

APPENDIX 4b

Data Processing System User Guidelines

SB2030 Data Processing System (DPS)
User Guidelines – Version 2.0
Updated: November 19, 1999
Prepared By: Larry L. Shannon, WRMA

PURPOSE OF SYSTEM

The SB2030 Data Processing System (SB2030DPS) has been designed to:

- Analyze the raw data which has been exported from the worker's personalized Excel Application;
- Clean and reduce the data files, removing those records not containing data;
- Provide some validation of the data;
- Monitor and report on the status of amounts of data being submitted; and
- Preparing the data for final analysis by via SPSS.

ENVIRONMENT – Software and Hardware

The SB2030DPS was written in Visual Basic 6.0 and runs on Windows 95/98/NT computers. The minimum desirable configuration is a Pentium 300 or higher with 128mb RAM, and 10-15gb of hard disk.

The SB2030DPS software is located in the C:\workload\system\ directory. It can be executed either in the Compiled Mode, using the SB2030.exe, via the SB2030 Icon, or in the Interpretive Mode, using the source code. It is highly recommended that the Interpretive Mode be used for a few days because of the additional ability to correct spurious errors that might occur during processing. Raw data of this type has the opportunity to be polluted easily by innocent users. In the Compiled mode, spurious errors will stop the SB2030DPS and prevent it from continuing.

REMOVING DATA FROM THE COLLECTION SERVER

Data is received from the Counties on the WRMA Collector Server, located in the State office building where Lee Stolmack works. The data are deposited on the Collector Server between midnight and 4:00am on a daily basis, including weekends. WRMA staff will collect the data from the Server on a daily basis (M-F). The data on the Server are in the format of zip files, residing under a Day of Year directory, e.g., 264 (the 264th day of 1999). This Julian Day number is set by the IBM Network Team. Each zip file represents the data swept from a single County Server within the Zone currently entering data. The filenames of the zip files are in the

form of Dccssvvv.zip, where D is any letter, cc is the County number, ss is the County Site number, and vvv is the County Server number.

The Collector Server resides within the offices of the IBM Global System Center and a "contractor" badge has been issued to WRMA and must be worn by the WRMA staff collecting the data. A new blank CD-ROM-R from WRMA must be taken to the State site. There is a procedure sheet from IBM that explains the tasks that must occur to collect the day's data. On Mondays, the Friday, Saturday, and Sunday data sets must be collected. All of the data for any daily collection will fit on one CD-ROM-R. After the data are written to the DC-ROM_R, the Day Directory is copied to the D:\drive and left there. The CD_ROM-R and its case must be clearly identified by the date of collection and the IBM Day number.

STORING THE SB2030 DATA

Upon return to the WRMA office, the staff must label the CD-ROM-R disk and save store it. One additional copy of the CD-ROM-R should be made and stored off site.

The labeling of the disks must contain the Day number. e.g., 264 and the date the data was collected form the Collection Server at the State office.

Currently Jeff Johnson and Vijay Mopuru have been designated to do the daily data collections.

SB2030DPS LAUNCHING AND SCREEN EXPLAINATION

Interpretive Mode Operation

1. Double click on the VB6PE icon to launch Visual Basic.
2. Select SB2030DPS from the File Menu, Currently used Programs, 1-SB2030. The SB2030DPS source code will load. There may be a load error mentioned. Click OK to it.
3. Launch SB2030DPS from the Run Menu, option Start. SB2030DPS will start. Proceed to use the program.
4. Close out the SB2030DPS from the File Menu, Exit.

Compile Mode Operation

Double click on the SB2030 icon to launch the program. Proceed to use the program. This mode of operation is not suggested for a week or so.

The Screen Controls

There is only one screen in the SB2030DPS. When the program is launched, this screen is in full view of the user. The controls on the screen will now be described, in general top left to bottom right.

Drive Selection Control – To select the data drive where the SB2030 data resides. (The SB2030DPS software must always reside on the C:\ drive).

Check Box, For Zip Files – Tells system to look only for *.zip files. This is the first process to be done on a given day's data, preparing for the upzip process by creating the County Server structure and placing the zip files in the proper location.

Check Box, For Data Files – Tells the system to look only for *.csv files. This is the third process to be done on a given day's data, collecting the data paths pointing to the workers' csv files.

Check Box, For Cross County Reports – Tells the system to process in the Cross County and Cross Day environment, e.g., preparing the Master Matrix of Server Activity.

Check Box, Write Error Records – Tells the system to write an image of the worker's input record to an error file if the record has been excluded during the processing.

Directory List Box – Displays directories for the user from which to select a given Day directory.

File List Box – (Upper Mid-screen) Displays a list of all servers (equivalent to a zip file) files from which the user should select the total set.

Display Field, Data Day # – Julian day assigned by the IBM Network Team for each separate data sweep.

Display Field, # CSV in Day – Total count of worker csv files being processed in the current Data Day.

Display Field, Processing CSV # - A running count of workers being processed in this Data Day.

Button, Get File List – Tells the system to select a set of files, either *.zip (if For Zip Files was checked by the user) or *.csv (if For Data Files was checked by the user).

Button, Prepare Unzip Batch File – Creates the DOS batch file (unzipper.bat) which, when executed, creates the proper directory structure and unzips all of the zip files into the proper directories.

Button, Process Day – Initiates the processing for a given Data Day.

Button, View Day Paths – Displays in the list box, below the button, the file paths to be processed in the Cross Reports (daypaths.txt file). This allows the user to review the days to be included in the Cross Reports.

File List Box – (Lower Right-screen) Displays a list of all servers (equivalent to a zip file) files from which the user should select the total set.

Button, County Reports – Initiates the Cross County/Day Reports process.

Status Line #1 – Indicates which key information about the currently running process, e.g., County and Worker ID csv file name currently being processed.

Status Line #2 – Displays the start and stop time of the process. (Not completely correct)

Button, Information – Not operational at this time.

Footer Field – Shows the WRMA software identification information, including the current version and date of the SB2030DPS system. Version 1.0 was the original version and Version 2.0 added all of the Cross Report features.

Button, Exit – Exits the system.

PROCESSING A DAY OF SB2030 DATA

It is essential the user follow these steps and process all of one Day at a time. DO NOT unzip many days, one after another and then process the Days. DO the entire process for one Day until it is completed.

Loading the Data into the Computer

Upon return to the WRMA office, the staff must copy the Day Directory from the CD-ROM-R DISK onto the processing PC at WRMA into the directory C:\workload\data\input\. The Day Directory must then be renamed to Date_Day to include a calendar date. The calendar date to be used, is the file date stamp on all of the Server Zip files for that Day. For example, if the Julian Day was 264 and the Server Zip file date was September 22, 1999, the C:\workload\data\input\264\ directory on the processing PC would be renamed to C:\workload\data\1999-09-22_264\.

Preparing the List of Raw Zipped Files

1 Launch SB2030DPS.
2 Select the proper drive.
3 Click in the For Zip Files Box. This tell the system to find all of the zip type files.
4 From the Directory List, click on the data directory containing the data for the Day to be processed, e.g., C:\workload\data\1999-09-22_264.
5 Select all of the files listed in the Files List Box, by clicking of the top one and scrolling to the bottom and Shift-Click on the last file.

6 Click on the Get File List button. The system will proceed to find all of the zip type files under the specified Day directory. Two OK displays will occur. The user should click OK for these.

7 Click on the Prepare Unzip Batch File button. This process takes the list of zip files from step 3 and prepares a DOS batch file which will 1) create a Server directory and then 2) unzip each Server zip file into an individual server directory. This operation ends by terminating the SB2030DPS because the user must now run the DOS Batch file (unzipper.bat) to upzip all of the files.

Unzipping the Files for the Day Directory

1 Click on the Windows Start button and select Run.

2 Choose the "drive":\unzipper command line. This drive letter must be the same as the drive selected by the user for the Data Day.

3 Click OK. The batch file unzipper.bat will execute. The screen wildly displays the batch file operation and all of the WinZip launches caused by the Unzipper command. There may be a hundred or more of these WinZip launches. Each of the WinZip processes will result in possibly hundreds of worker *.csv (comma separated value) files being unzipped into the Server directory.

4 Wait for all of the screen action to stop.

5 If a WinZip operation terminated with no problems, it will close the screen and disappear by itself.

6 If a WinZip operation does not terminate properly or has problems, the red light at the bottom right ofd the screen will be on and the user must take some action. Usually there will be a dialog box of some type to indicate the problems found.

7 Resolving WinZip issues. The following dialog boxes may appear: a) some Internal Count (byte count) is wrong (Click OK, the user cannot solve this. b) Some non zip file was found and the user can only try to get past it (Click OK). c) Do you want to resume processing the rest of the zip file? (Click YES) Usually only one file has been lost (there are 2-3 of these occurring per Day now) d) Do you want to save the WinZip error information? (Click NO) We may look into these later.

8 If a WinZip screen is on top and the green light at the lower right corner of the screen is on, the process had problems, but the user has satisfied the issues. Click on the "X" in the upper right corner to close the green light WinZip process.

9 When all of the WinZip processes have been cleared out, the Batch operation remains. The contents of the DOS window will be static. Click on the "X" in the upper right corner to close the DOS unzipper.bat program.

Preparing the List of Raw CSV Data Files

1 Launch SB2030DPS.

2 Select the proper drive.

3 Click in the For Data Files Check Box. This tell the system to find all of the csv type files under the Day directory.

4 From the Directory List, click on the data directory containing the data for the Day to be processed, e.g., C:\workload\data\1999-09-22_264. The Files List Box will remain empty.
5 Click on the Get File List button. The system will proceed to find all of the csv type files under the specified Day directory.

Starting a Data Day Processing

1 Click on the Process Day Button. This will initiate the full processing of the Data Day. The Data Day #, # CSV in Day, and Processing CSV # fields will be displayed.
2 At the completion of the processing, click on Exit to terminate the system.

Description and Location of Data Day Input Files

The following files are input to the Data Day Processing:
1 "drive":\workload\data\input\yyyy-mm-dd_jjj\"server"\cc_www.csv – Worker ASCII data files from the Excel Application from one County Server.

Description and Location of Data Day Output Files

The following file are written as a result of the Data Day Processing:
1 "drive":\workload\data\output\Short_server.txt – A quick list of servers reporting from the various Counties for the Data Day. Totals are also included by County. This report is an early report to be distributed after the unzipping, but before the full processing for the Day.
2 "drive":\workload\data\output\Server_paths.txt – List of drive paths pointing to all of the server zip files.
3 "drive":\workload\data\output\Unzipper.bat – This is the DOS batch program that creates the structure for the Input directory and launches each of the WinZip applications to unzip the server files into worker csv files.
4 "drive":\workload\data\output\Server_stats.txt – Excel compatible list of the detailed stats on the servers reporting for the Data Day. This report is not finished until after the total processing has taken place.
5 "drive":\workload\data\output\Short_Server_Participant.txt - A quick list of participants within the servers reporting from the various Counties for the Data Day. Totals are also included by server. This report is an early report to be distributed after the unzipping, but before the full processing for the Day.
6 "drive":\workload\data\output\Particip_paths.txt - List of drive paths pointing to all of the participant csv data files.
7 "drive":\workload\data\output\ErrorRec.txt – Contains record images from input transactions that were rejected. Only active if the Write Error Records checkbox has been activated.
8 "drive":\workload\data\output\Charact.txt – Contains all of the Characteristic records from the Data Day.

9 "drive":\workload\data\output\DayTabs.txt - Contains all of the Day Tab records from the Data Day.
10 "drive":\workload\data\output\Cover.txt - Contains all of the Cover Sheet records from the Data Day.
11 "drive":\workload\data\output\ZipSurvive.txt – Report indicating whether an incoming zip file actually survived and produced good data, e.g., was the zip file empty, not a zip, etc.
12 "drive":\workload\data\output\WorkerStat.txt
13 "drive":\workload\data\output\Particip_stats.txt – Excel compatible list of the detailed stats on the participants for the Data Day. This report is not finished until after the total processing has taken place.
14 "drive":\workload\data\output\Caseload.txt - Contains all of the Caseload records from the Data Day.
15 "drive":\workload\data\output\Master_roll.txt – A list of all of the servers in the State with information about how many reported for the current Data Day.
16 "drive":\workload\data\output\C"countyID"par.txt – Identical to the Particip_stats.txt file, but containing only data from the given County.
17 "drive":\workload\data\output\C"countyID"svr.txt – Identical to the Server_stats.txt file, but containing only data from the given County.

Description and Location of Cross Report Output Files

The following file are written as a result of the Cross Reports Processing:
1 "drive":\workload\data\Cross_reports\DayPaths.txt – List of all Date-Day directory paths which are to be included in the Cross Report processing. This is automatically updated if all of the processing is done on one PC. Otherwise, the user must keep it up to date.
2 "drive":\workload\data\Cross_reports\Critical_server_hist.txt – This is a report indicating in English various warnings about the status of the server activity history.
3 "drive":\workload\data\Cross_reports\Master_roll_hist.txt – Identical to the master_roll.txt file but for cross Counties.
4 "drive":\workload\data\Cross_reports\Particip_stats_hist.txt – Identical to particip_stats.txt, but across days and Counties.
5 "drive":\workload\data\Cross_reports\Server_Matrix_Hist.txt – Master chart of server activity history across all days and all Counties.
6 "drive":\workload\data\Cross_reports\Server_stats_hist.txt – Identical to server_stats.txt, but across days and Counties.
7 "drive":\workload\data\Cross_reports\C"countyID"par_Hist.txt - Identical to the Particip_ stats .txt file, but containing cross days data from the given County.
8 "drive":\workload\data\Cross_reports\C"countyID"svr_Hist.txt - Identical to the Server_ stats .txt file, but containing cross days data from the given County.
9 "drive":\workload\data\Cross_reports\C"countyID"svr_Mtx.txt – Identical to the Server_ Matrix_ Hist.txt file, but containing cross days data from the given County.

Running Cross Day and Cross County Reports

After the processing for the Day is done, cross reports may be produced by the user. There is one file that is semi-automated for these cross day; and cross County processes. The system keeps this file up to date, but, the user needs to review them each time before executing any of the reports. As indicated above, if processing is done on multiple computers, the file will not reflect the correct paths. The file is:

> "drive":\workload\data\cross_reports\output\daypaths.txt – This file contains paths to all of the Data Day output directories to be included in the Cross Reports processing. Each time a Day is processed, another record is added to this file. The file contains records like:
>
> "drive:\workload\data\output\1999-09-22_264\

Once the user has verified, using the Notepad or Word, that these two files are up to date, the reports can be processed. (Or see the View Day Paths button.) If the reports are to cover only one Day, there should only be one record in this file. No two records in this file should match. If selective output paths are to be processed, the user may place a single quote (as a comment line) at the front of all records not desired to be included. To start the report process, the user must click on the County Reports button.

During the report processing, several messages will be displayed for the user. The user should click on the appropriate button, usually OK.

Dissemination of Reports

The following reports are distributed on a regular basis, via e-mail:

To SB2030 Workload Team:

> "drive":\workload\data\Cross_reports\Server_Matrix_Hist.txt – To indicate the history of the reporting servers across all of the Counties and days.
>
> "drive":\workload\data\Cross_reports\C"countyID"par_Hist.txt – To indicate the participant history across all of the Counties and days.

To the IBM Network Team:

> "drive":\workload\data\Cross_reports\Server_Matrix_Hist.txt – To indicate the history of the reporting servers across all of the Counties and days.

To the Counties:

> "drive":\workload\data\Cross_reports\C"countyID"svr.txt – To indicate the County's server activity for a given Data Day.

"drive":\workload\data\Cross_reports\C"countyID"par.txt – To indicate the County participant activity for a given Data Day.

Revised Content of the Particip_Stats.txt File (Version 11-16-99)

The following fields have been appended to the end of the Particip_Stats.txt file in the output\1999-mm-dd_nnn\ directory.

CSVName - Full csv filename

CSVDate - CSV File date Stamp

FileKept - quasi-indicator if the file was kept. (not currently meaningful)

RecsIn - Number of Records input

RecsKeep - Number of records kept (all types)

RecsErr - Number of records in error (all types)

CovOut - Number of Cover records output

CovErr - Number of Cover records in error

CaseOut - Number of Caseload records output

CaseErr - Number of caseload records in error

CharOut - Number of Characteristic records output

CharErr - Number of Characteristic records in error

Q(1-14) - Quality Flag for this Day (=1 if there is at least one record in the day, else =0)

QualSum - Sum of the 14 Quality flags (how many days reported with quality data)

H(1-14) - Total hours by the day

HrSum - Sum of the total hours for the csv file

DayTxsOut - Number of total day transactions output

DayTxsErr - Number of total day transaction error records detected

D(1-14)Out - Count of the transactions output by the day

D(1-14)E(1-5) - Count of the types of errors (1-5) detected for the transaction records, where the error types are:

SB2030 Data Processing Error Messages

01-Any Day Tab and Record Number not numeric
02-Any Day Tab, Service Category + Service Unit + Next Field = blank
03-# sign in Computed Time Field. Set to one #.
04-Illegal County Length. Use first two characters.