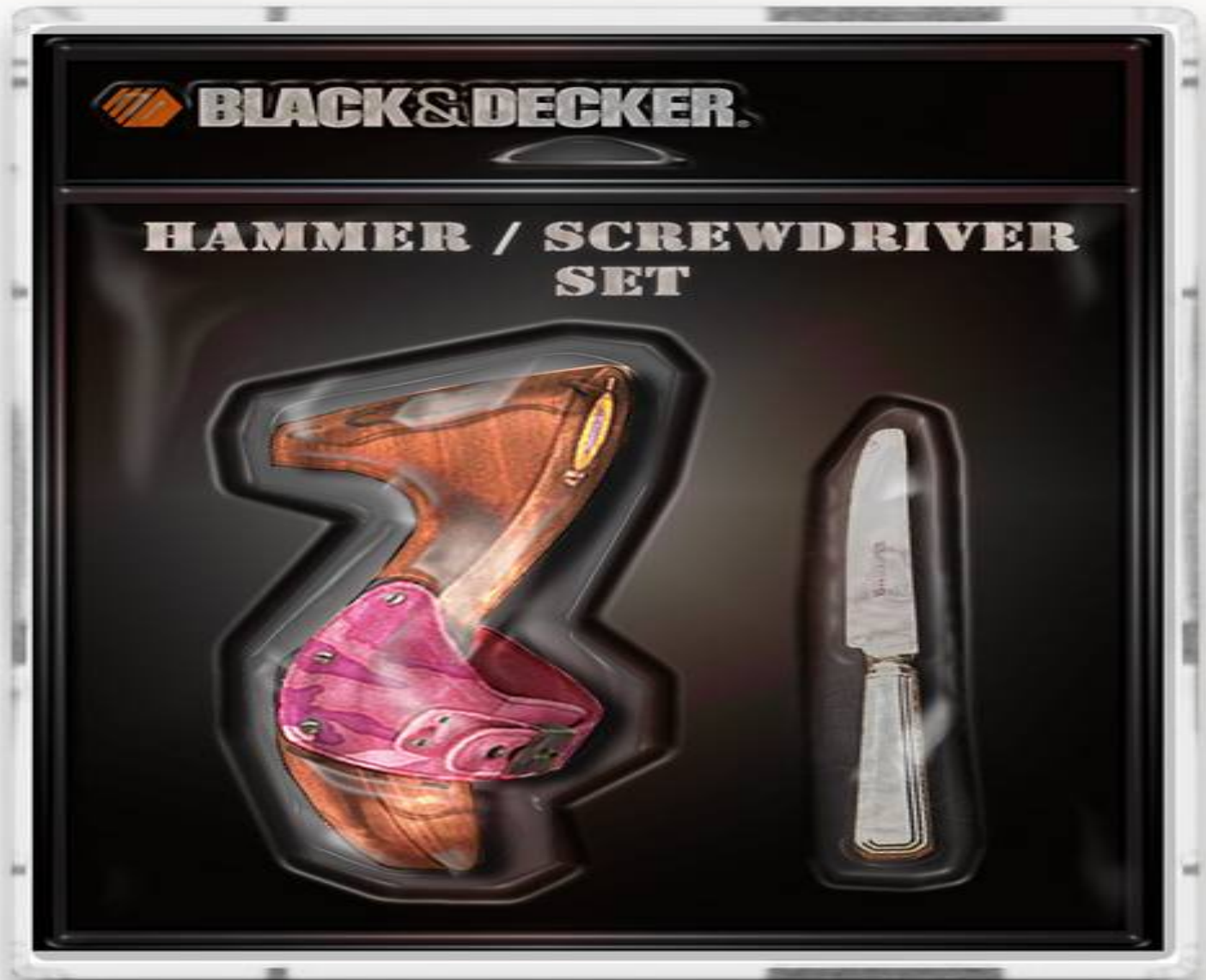


Unique Identifiers within Washington State's Department of Social and Health Services

NASMHPD Research Institute
Technical Assistance

April 3 & 4, 2006

It's important to have the right tools



Barriers to Integrated Data

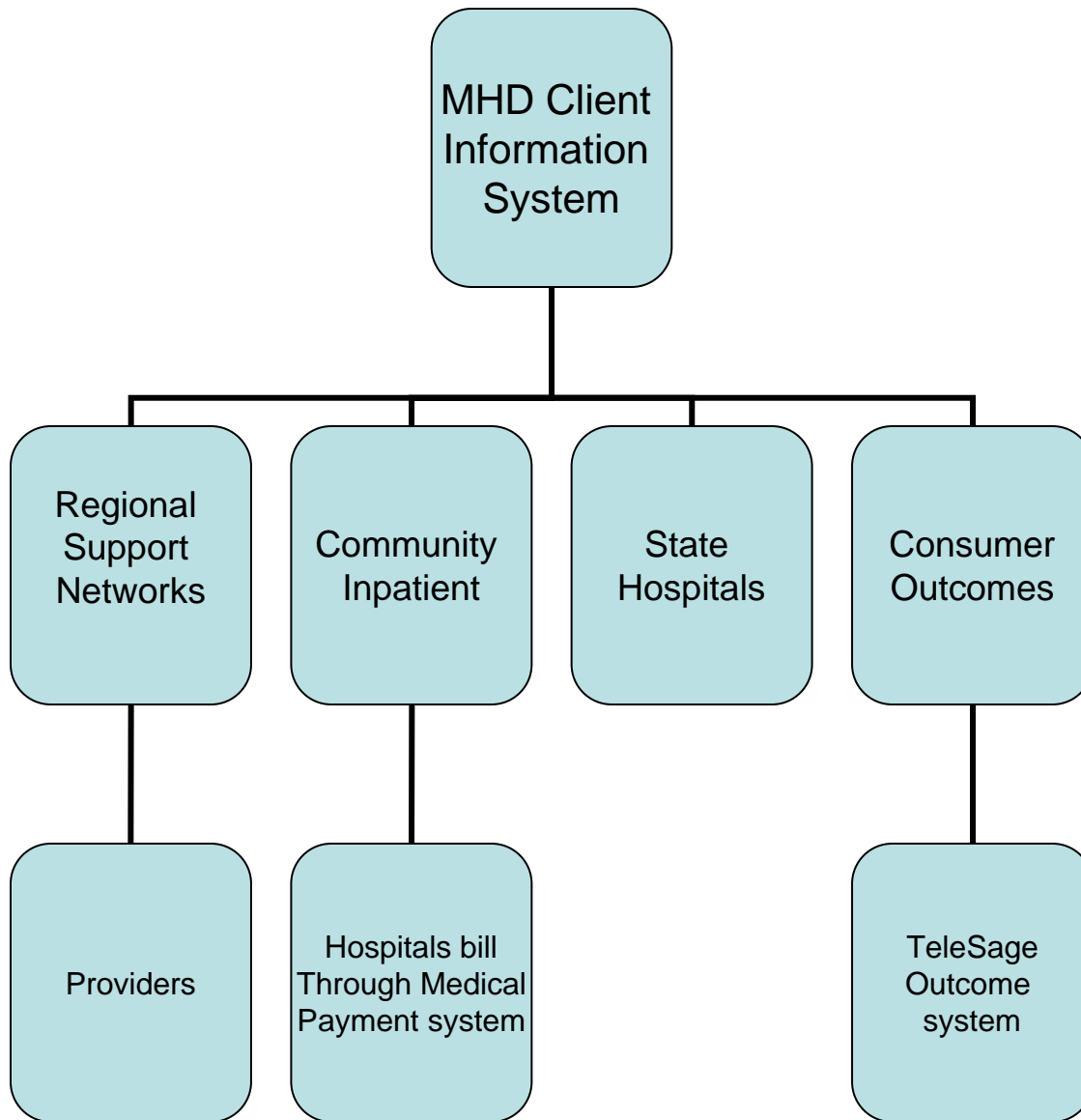
-Borrowed from Tracy Leeper, OK

- 42 CFR
- HIPAA
- IRB Approval
- Getting Buy-in from Administrators
- Creating On-going Working Relationships with Agency IT staff

Unduplication Methods

- Two types of unduplication methods used in Washington
 - Multi-level match-Merge
 - Deterministic Linking
- Multilevel used within MHD
- Deterministic used across DSHS

Mental Health Division Data Map



- Each level uses a different client ID
- Clients unduplicated in the MHD-CIS

Data dictionary, as part of contract, requires RSNs to submit data in a specified format (837P)

© 2000 Randy Glasbergen.
www.glasbergen.com



"We rarely back up our data. We'd rather not keep a permanent record of everything that goes wrong around here!"

Washington Mental Health Division Matching Algorithm

- 5 data elements are used to unduplicate client records
 - Given name (first and last)
 - Date of Birth
 - Social Security Number
 - Gender

Washington Mental Health Division

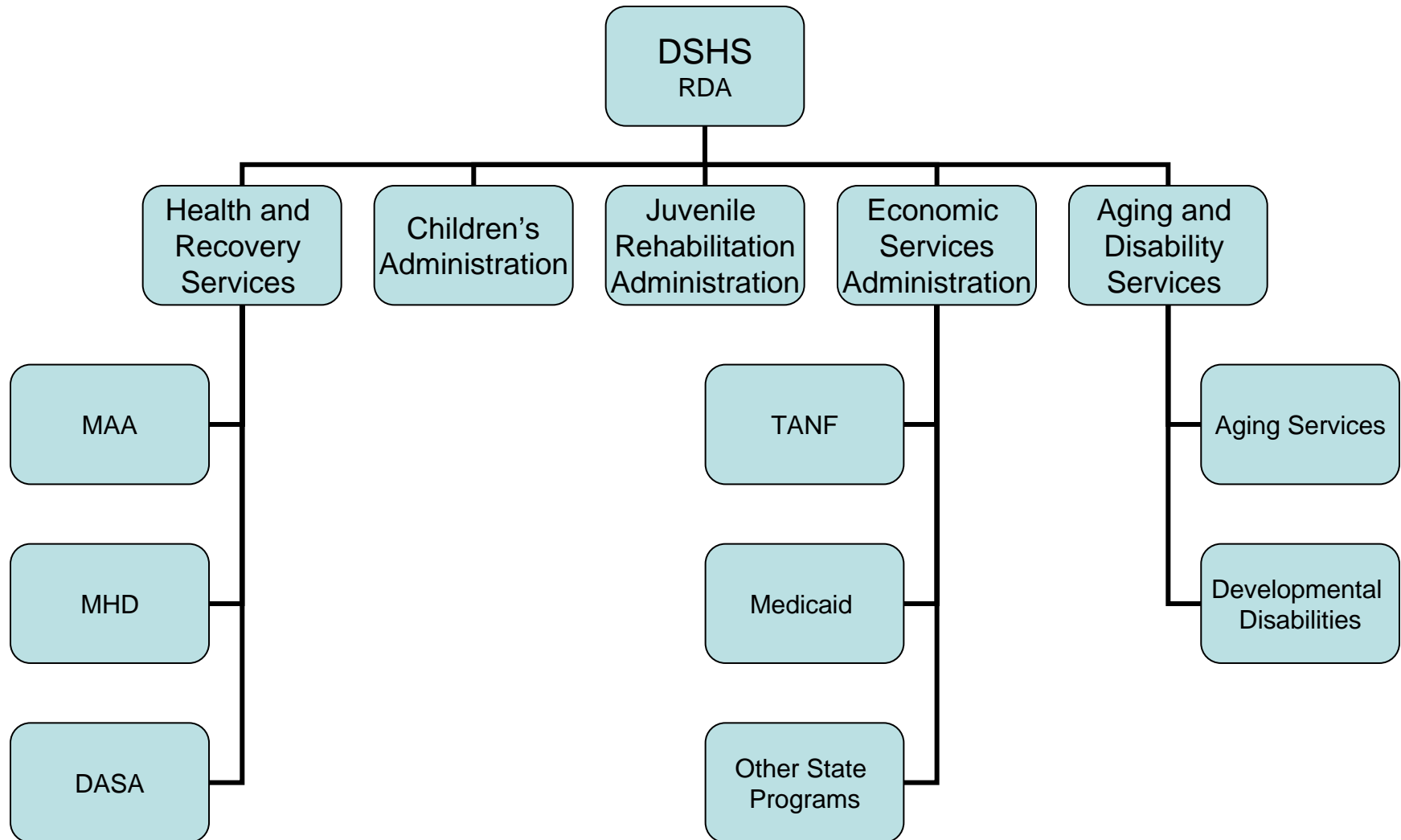
Matching Algorithm

- **First Match – Name, DOB, Gender**
 - Must have valid date for DOB, and surname is not “unknown”, “Doe” or blank.
 - A match exists when last name, first name, DOB and gender are the same - or last name, DOB and gender are the same
- **Second Match – Last name, SSN**
 - Must have valid SSN, and last name is not “unknown”, “Doe” or blank
 - A match exists when last name and SSN are the same
- **Third Match – SSN, DOB**
 - Must have a valid DOB, SSN
 - A match occurs when DOB and SSN are the same.

MHD Identifier

- Unique Identifier is computer generated
 - Contains no client specific information
 - Updated each time new information is received
- Data dictionary, as part of contract, requires RSNs to submit data in a specified format (837P)
- IT staff work with Medical billing and state hospitals re: changes made to data formats.

Department of Social and Health Services Data Map



DSHS Matching Algorithm

- 5 data elements are used to unduplicate client records for the Client Services Data Base

weights

- ACES Eligibility number (14.1)
- Given name (first , middle) (6.1)
- Family Name (last) (5.0)
- Date of Birth (6.0)
 - Program checks for inversion of month and day
- Social Security Number (8.2)

DSHS Matching Algorithm

- A match on each variable is given a weight.
 - Weights were derived from matching analysis
- Records with weights greater than 14.15 are considered matches and unduplicated.
- Unique Identifier is computer generated
 - Contains no client specific information
 - Updated quarterly

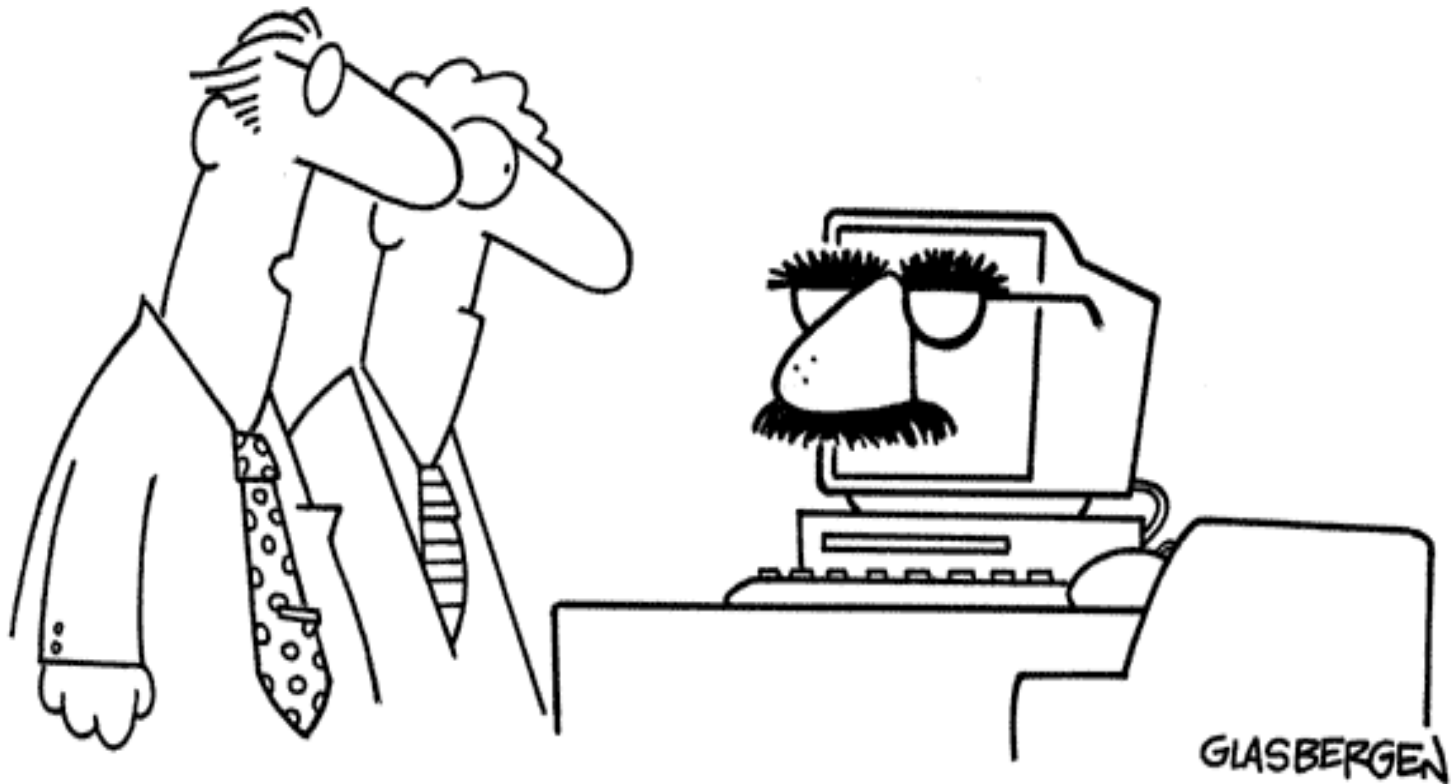
Client Services Data Base

- A team of RDA staff work with agencies and programs to track changes made to data formats.
- Minimal data is maintained from each agency
 - Client identifiers
 - Service hours
 - Service costs

Limitations

- Data additions are cumbersome
- Data are 18-24 months old at any given time.
- Multiple Research data bases have been created that combine fewer agencies, include more detail, and are closer to “real-time”.

Copyright 2004 by Randy Glasbergen.
www.glasbergen.com



“I’m sure there are better ways to disguise sensitive information, but we don’t have a big budget.”

Confidentiality

- All programs have access to aggregate data reports
 - www1.dshs.wa.gov/rda/research/clientdata/2004/default.shtm
- Only RDA staff have access to identified client data
- For programs to get access to identified client data:
 - If research, need IRB approval and data sharing agreement
 - If treatment, planning, operations or program evaluation, need data sharing agreement
(many programs also have reviewed by IRB)

Conclusions

- Integrated databases are possible and very useful
- They require:
 - a multi-year commitment
 - long development and start-up period
 - dedicated staff and IT resources
 - cooperation across agencies
 - patience