

The Processes CNEAF Follows To Assure Data Quality: The Case of Coal *

**ASA/EIA Meeting
October 19, 2007**

**William Watson
Office of Coal, Nuclear, Electric & Alternate Fuels**

* This is a working document prepared by the Energy Information Administration (EIA) in order to solicit advice and comment on statistical matters from the American Statistical Association Committee on Energy Statistics. This topic will be discussed at EIA's fall 2007 meeting with the Committee to be held October 18 and 19, 2007."



IDC Deployment

<u>Survey</u>	<u>Frame Size</u>	<u>Population</u>	<u>Frequency</u>	<u>Deployed</u>	<u>Cycle</u>
EIA-6A	960	Distributors	Annually	12/04	2005
EIA-7A	1,540	Coal Mines	Annually	12/03	2003
EIA-3	2,000	Mnfg. Plants	Quarterly	03/04	2004 Q1
EIA-5	80	Coke Plants	Quarterly	03/04	2004 Q1



Survey Response Rates

<u>Survey</u>	<u>Period</u>	<u>% of Frame</u>	<u>% IDC</u>	<u>% of Current Year Production</u>
EIA-7A	2002	-	0%	-
	2003	-	55%	-
	2004	90%	63%	98%
	2005	94%	75%	99%
	2006	94%	74%	99%
EIA-6A	2004	94%	0%	-
	2005	94%	53%	-
	2006	99%	64%	-
EIA-3	All Qs since 2004	100%	>95%	-



Frame Design & Maintenance

- Specify threshold
- Establish a base frame
- Revise frame
 - Births
 - Deaths

Data Survey Forms

- Instructions
- User-friendly interface and flow to simplify process and minimize human error
 - For example, drop-down lists, standardized names
- Automated accounting checks
 - Respondents resolve imbalances
- Automated edit flag checks
 - Respondents resolve invalid data edit flags by entering correct value or an override comment



Respondent Contact Record Tool (RCRT)

- Tracks unresolved edits & their resolution
- All forms with data edit flags are given a default “Pending” status
- Includes every IDC form submitted with an override comment
- Contractors review comment, perform “Fast Acceptance” or follow up with respondent by phone or e-mail
- Federal staff review “Fast Acceptance” cases and identify forms for recheck
- Survey closed when all RCRT statuses are “Completed”



Level 3 Checks

- Survey data sorted by specific fields from lowest to highest value to identify outliers
- Example: Respondent reports 0.04 as sulfur value, value fails edit range check of $1 < S < 5$, respondent enters override comment: “value as reported by lab,” contractor and Federal staff accept comment as valid. The value should have been entered as 4, representing 4 percent.



Level 3 Checks

- Data survey fields searched for “null” entries
- Page 1 – page 2 comparables checked
 - Page 1 has total receipts
 - Page 2 has receipts at detailed level
 - Sum of page 2 receipts should equal page 1 receipts

QA Guidance Documents & Process

- Checklist of all process steps, QA, etc.
 - Track progress of survey
 - Document issues & problems
 - Recommend changes
 - Sign-off at end of reporting cycle
 - Submit to CNEAF QA officer



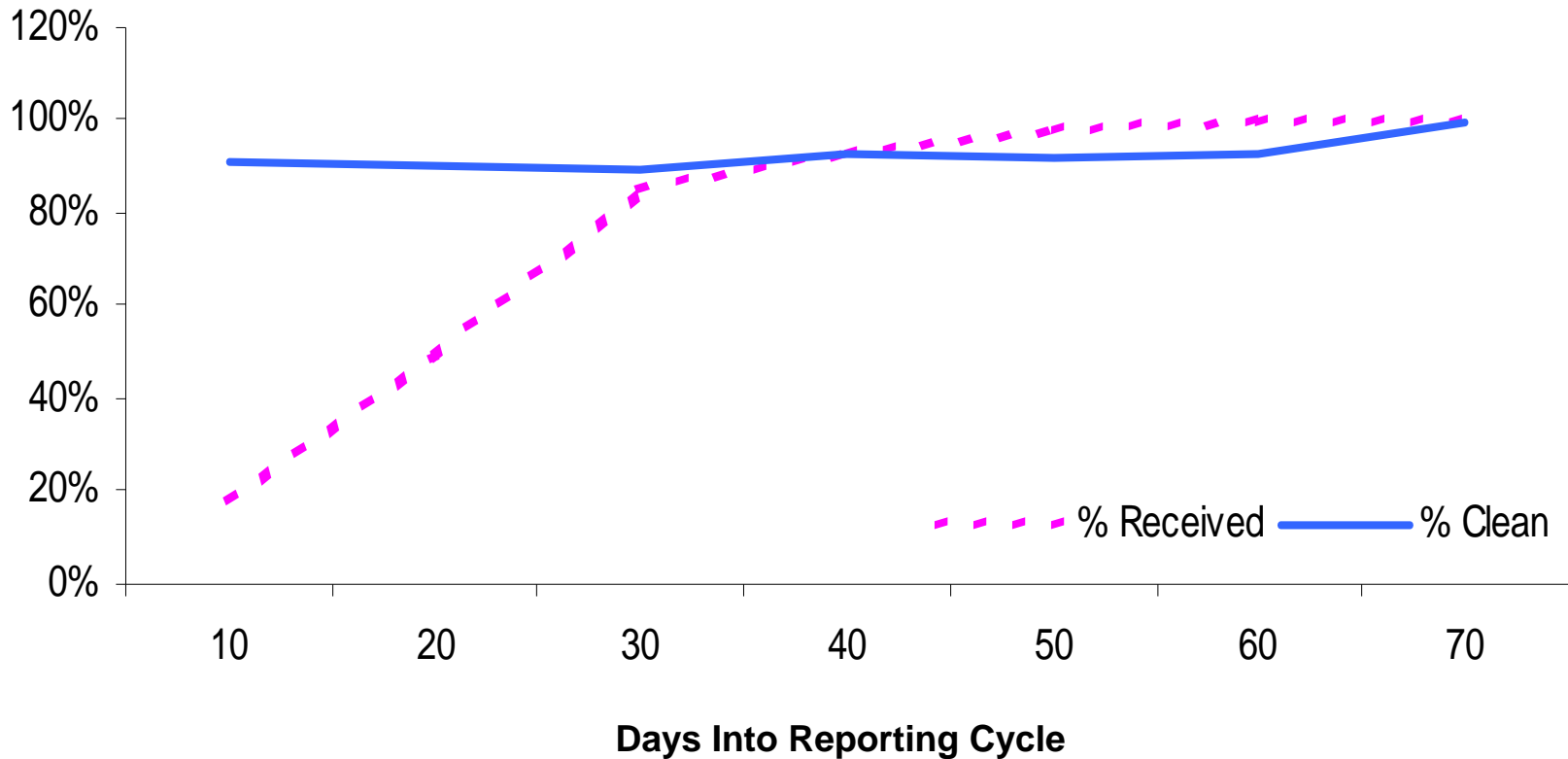
QA Guidance Documents & Process

- Weekly or bi-weekly status review
 - Response rate charts
 - Clean-up progress charts
- End-of-cycle wrap-up review
 - Review of what worked
 - Review of what didn't work
 - Review of what needs fine tuning



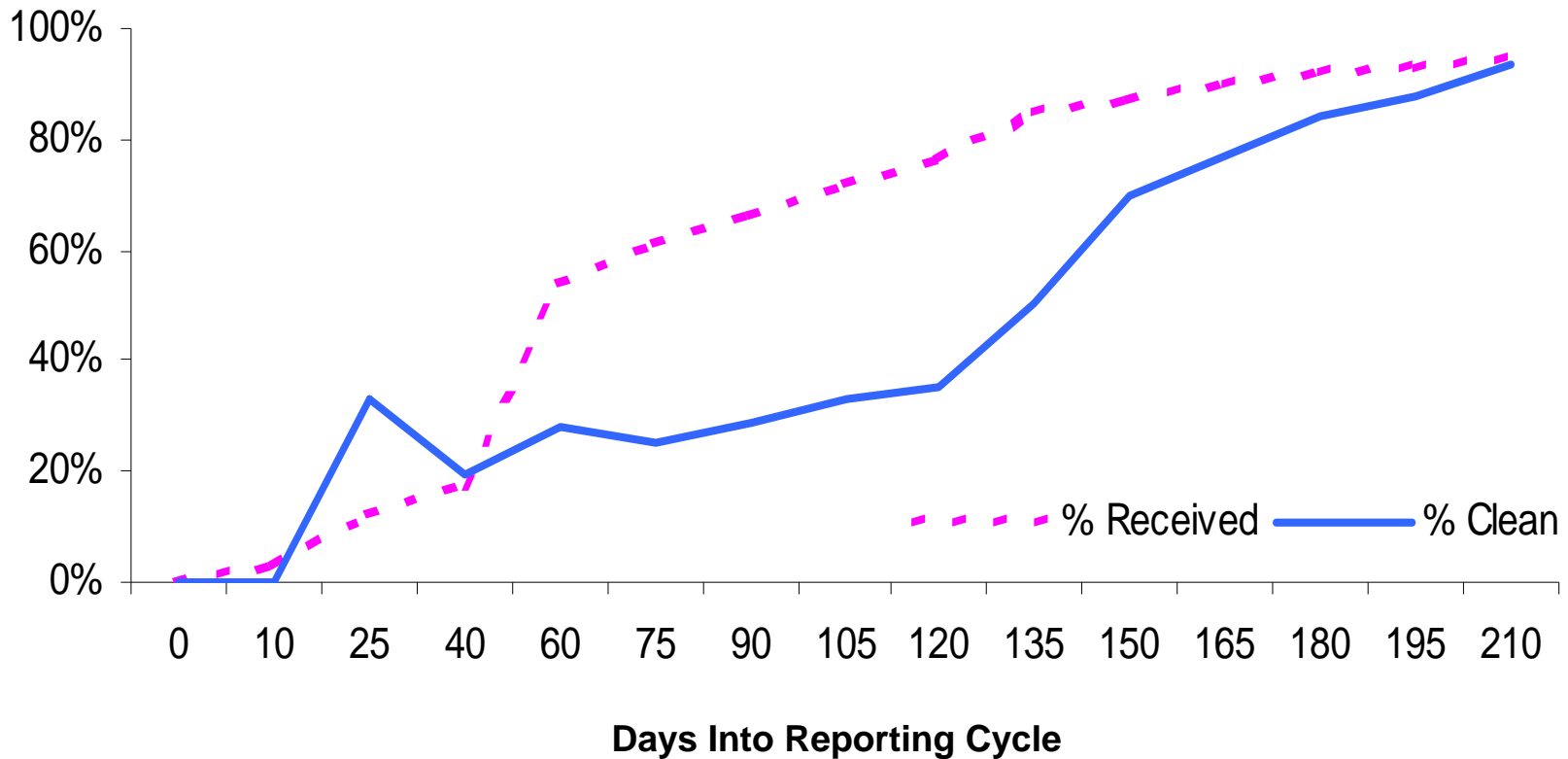
EIA-3 Responses

EIA-3 Manufacturing Survey 2007 Q1

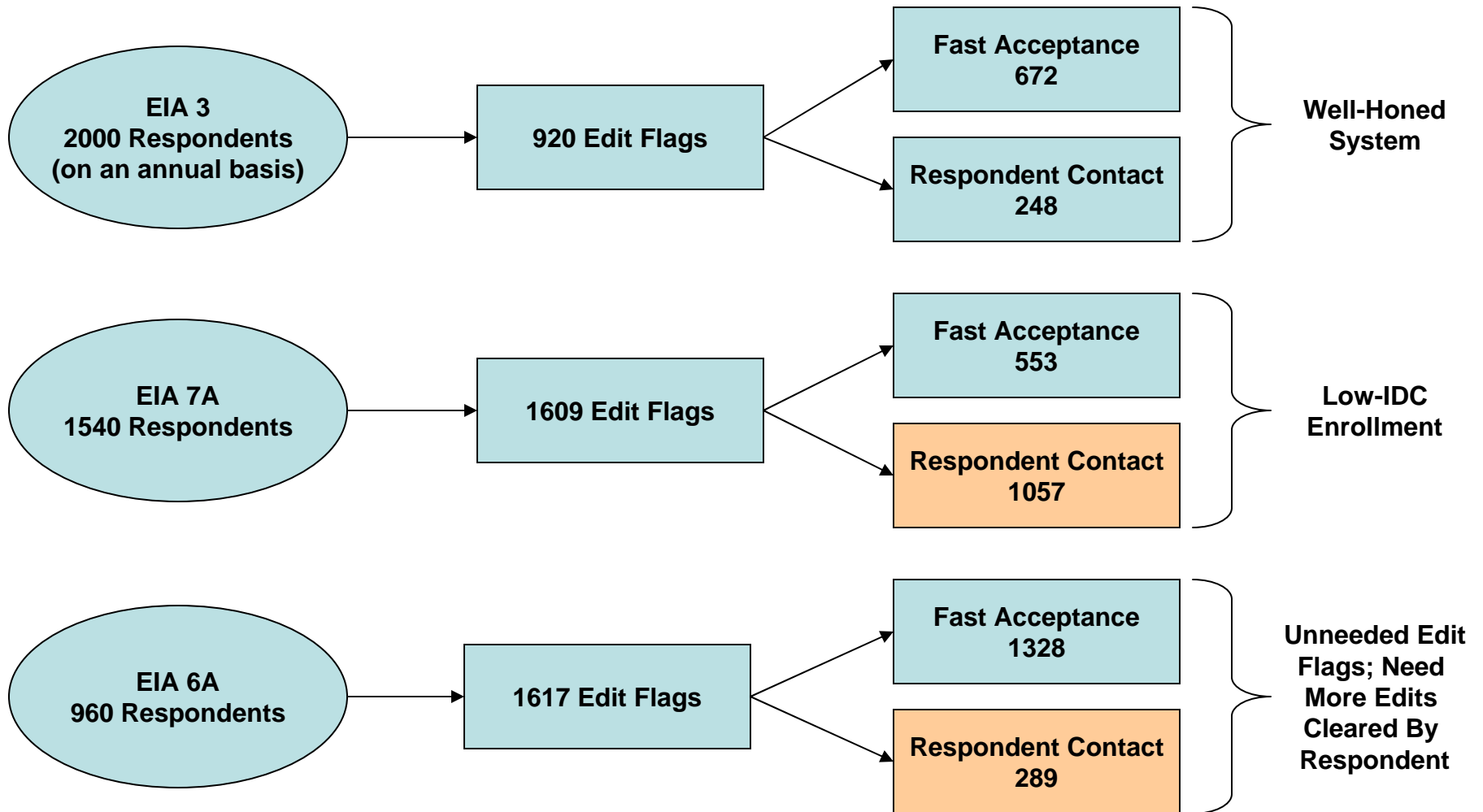


EIA-7A Responses

EIA-7A Coal Mine Survey 2006



Edit Resolution in Practice



Micro Strategies to Improve Quality

- IDC enrollment > 95%
- Self-editing by respondent
 - Complete and balance data to progress through form
 - Data edits against past values and industry standards
- Fast compilations of edit flag status
- Fast review/resolution of pending edits
- Cross survey comparison



Macro Strategies to Improve Quality

- Frame maintenance – EIA3
 - Industry newsletters and reports (new coal users)
 - Inventories to implement environmental regulations
 - Cross reference to other surveys
- Aggregate Balance
 - Production + Imports + Stock Change = Consumption + Exports
 - OR
 - Supply = Demand
- Challenge: Supply routinely exceeds demand

Questions to ASA Committee

- What weaknesses do you see in our QA process?
- What can we learn about data survey methods from academic sources?
- Are there best practices we should know and follow?
- How can we assure continued quality improvement?

