

**CRANE CENTER FOR
EARLY CHILDHOOD
RESEARCH & POLICY**

Managing Data Over the Lifecycle of Grant

Quality Assurance Practices

Kimberly Murphy
PhD Candidate



THE OHIO STATE UNIVERSITY

My Background

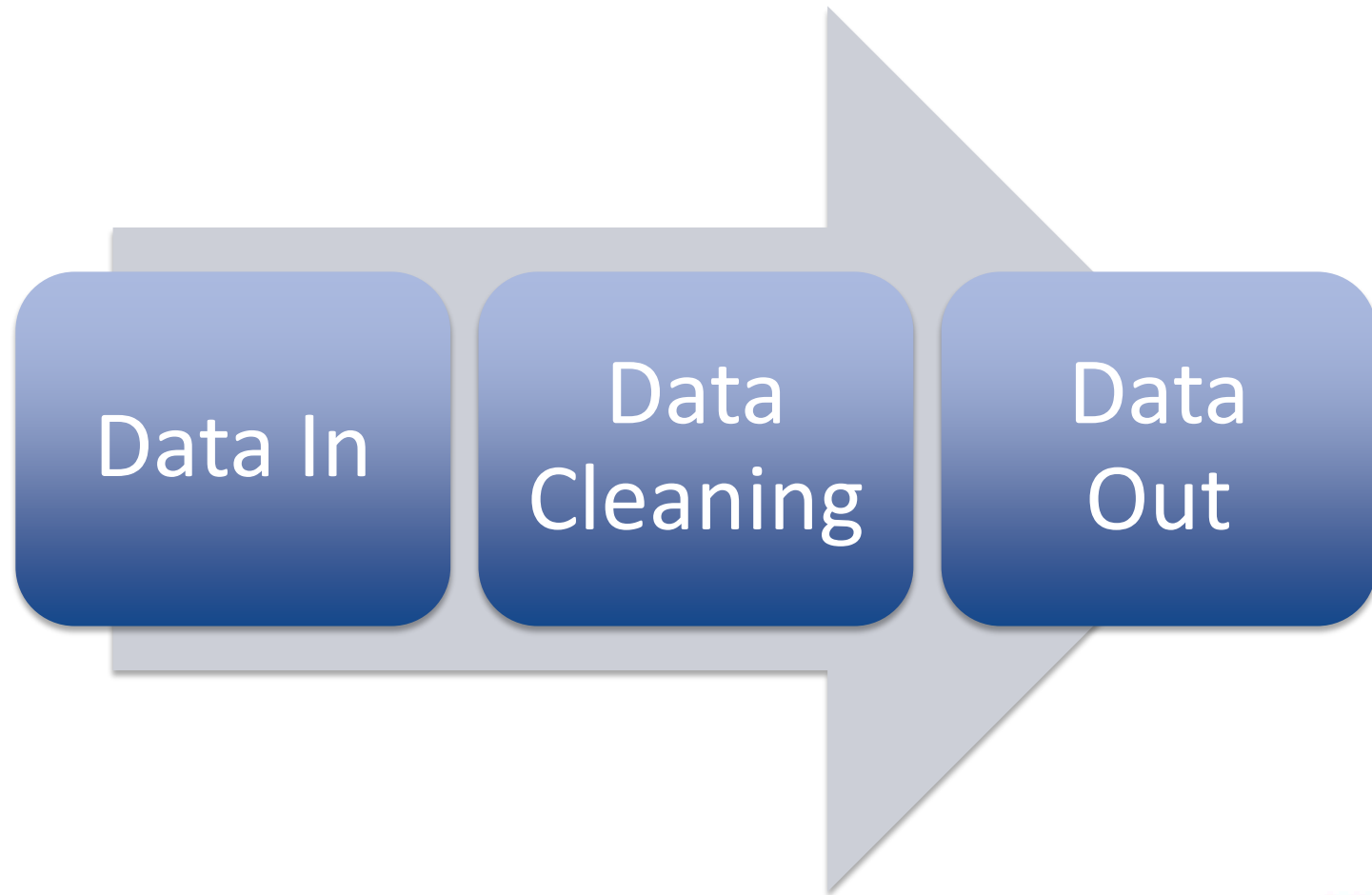
I work as a Graduate Research Associate for the Language and Reading Research Consortium (LARRC).

LARRC Assessment Study:

- multi-site 5-year research study
- longitudinal cohort design
- direct assessments, questionnaires, classroom observations



Data Management



Data Management: Recommended Practices

<http://onlinelibrary.wiley.com/doi/10.1111/j.1540-5834.2006.00402.x/abstract>



THE OHIO STATE UNIVERSITY

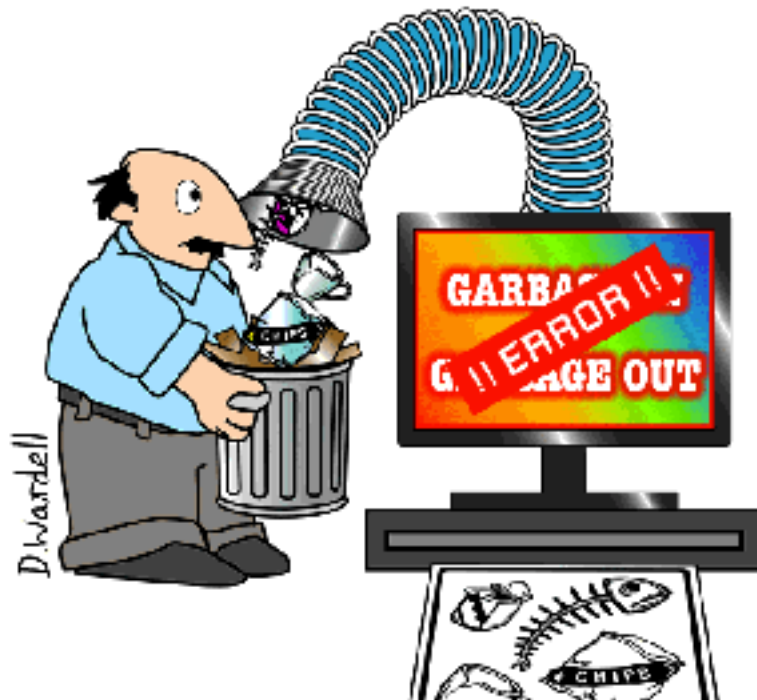
Data In

- Planning
- Data collection
- Data entry



Planning

Planning for data collection that will go smoothly and lead to data that is as clean, or as easy to clean, as possible - minimize errors later in the cleaning process



Don't put garbage in!

Planning

- Choose established, reliable measures (ideally)
- Method of data entry/capture (e.g., TeleForm, Access, Qualtrics) – use forms that are prepopulated with key info such as IDs
- Type of database and software (e.g., SQL, SPSS)
- Method of scoring (e.g., by hand in the field, postscoring, computer code)



Planning

- Develop protocols – for everything!
 - Assigning of participant IDs
 - Tracking changes in participants
 - Variable naming
 - Missing data rules and codes
 - Administration and scoring of measures
 - Basal and ceiling rules
 - Assessor training and recalibration/drift checks
 - Tracking data entry (especially if multiple sites)
 - Data cleaning
 - Error logs



Planning

- Codebooks
- Data dictionary

- Communication between all involved parties – researchers, assessors, tech support, RAs, etc.



Planning

- Assessor training
 - Develop clear, detailed training materials. Test them out.
 - Implement training:
 - materials: admin protocol, scoring protocol, PowerPoint, video, quiz
 - mock administrations
 - fidelity check
 - Research on assessment fidelity – e.g., see Reed et al., 2013 and 2014



Data Collection

- **ACCURACY**
 - Monitor the data collection, scoring accuracy & reliability, etc.
 - Recalibrate assessors, especially if data collection occurs over a long time period
- Minimize missing data



Data Entry

- **ACCURACY**
 - Check data forms for errors before entry
 - Use a double-entry or verification process
- Keep track of what is entered
- Minimize missing data

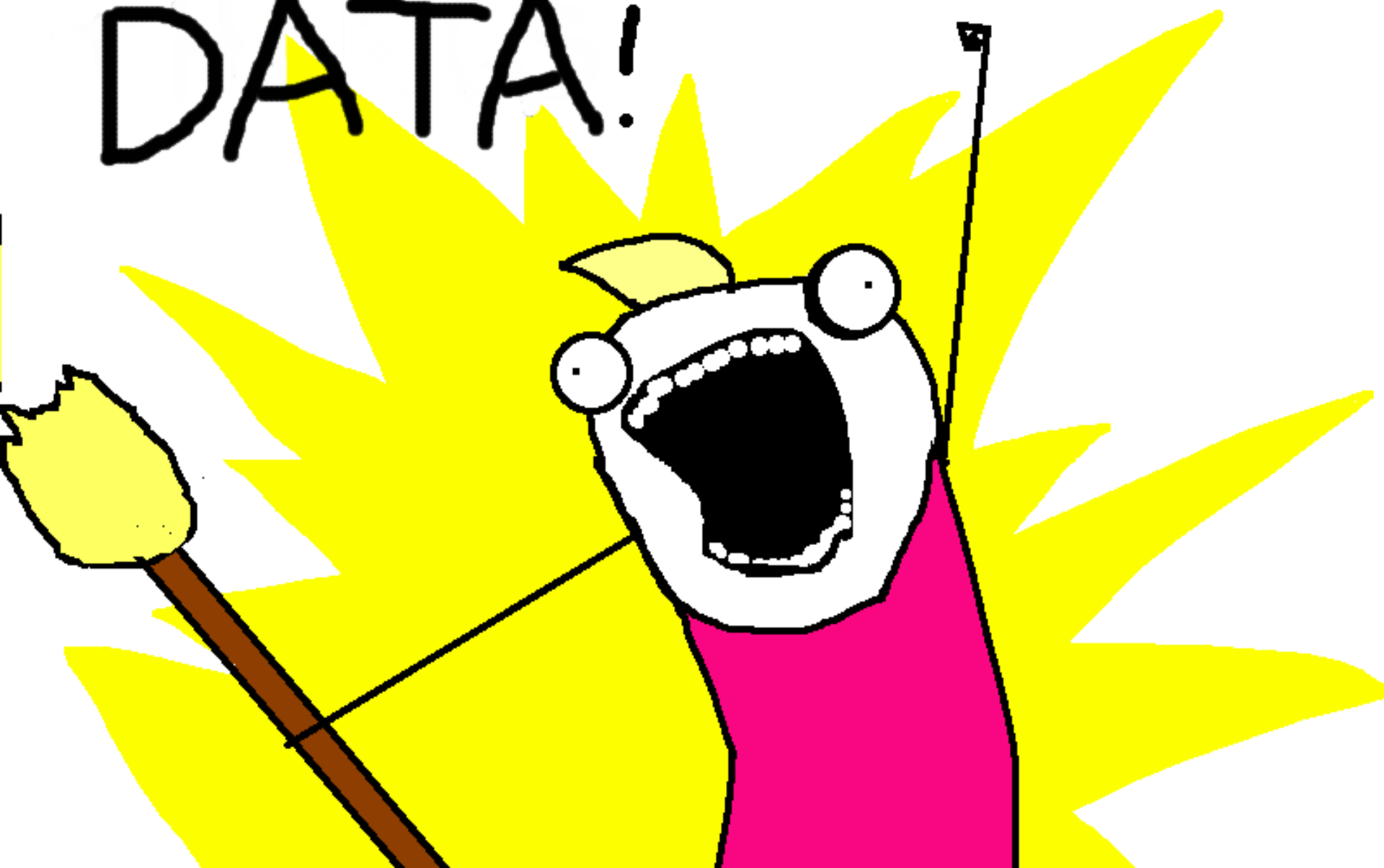


Data Cleaning

- Please don't skip this step 😊
- Clean data is a necessary element of high-quality data
- Will save you time (and frustration) later when you go to analyze your data



CLEAN ALL THE
DATA!



Data Cleaning

- Cardinal Rule: know your data!
 - What are the variables? the relationships between variables?
 - Possible ranges, rules for basals and ceilings, etc. for each measure
- Save your original file and separate versions used in data cleaning
- Be consistent in decisions re: handling errors
- Keep a log of changes, decisions, etc. in codebook



Data Cleaning

- Check for errors at every step in the process
 - Data collection (e.g., inaccurate test administration)
 - Data entry (e.g., errors in getting the data into your database)
 - Scoring (e.g., errors in calculations or applying basal/ceiling rules and missing data rules)
- Look at original forms and compare to database
- Check for duplicates
- Confirm ID list



Data Cleaning

- Looking at the numbers:
 - Run descriptives, frequencies, distributions
 - Full sample and subsets (e.g., age, grade, site)
 - Check outliers, valid ranges & values, calculations
 - Check correct use of administration rules such as basal/ceiling, reverse scoring, etc.
 - Look for unexpected patterns
 - Missing data



Data Cleaning

- A helpful resource for using SAS to check your data:
<http://www.ats.ucla.edu/stat/sas/library/nesug99/ss123.pdf>



Data Cleaning

- A broader view:
 - consistency and relations across measures – e.g., correlations between related measures
 - and across timepoints – e.g., compare descriptives
- Iterative process
 - Identify error – fix it – check it, etc.
- Final review of dataset



Data Out

- Preparing to analyze
- Sharing
- Set up the datasets to be user-friendly
 - Well organized, with values and labels included
 - Provide merged datasets
 - Codebooks, data dictionary
- Accessibility: Storage & archiving
- Security



A decorative header with a repeating pattern of stylized leaves and flowers in shades of gray on a white background.

Thank you!

For more information, contact me at
murphy.1116@osu.edu

My work at the CCEC is supported by
The Language and Reading Research Consortium
IES Grant # R305F100002