

QA/QC: Reducing Error Entry/Finding and Correcting Errors

Tools

Reduce Data Entry Errors with 'Data Validation' Rules

1. Data Validation in Excel [[link](#)]

Confirm data values after entry with 'Text-to-Speech'

1. MS Excel (& other MS programs)
 - Converting text to speech in Excel for Microsoft 365, Excel 2010-2021 [[link](#)].
 - Microsoft for Mac: Hear selected text read aloud from Excel, Word, Power Point, and Outlook [[link](#)]. See also '*System Preferences->Accessibility*'
 - Video Tutorials
 1. [Tutorial Video 1](#). This will allow you to select rows or columns of Excel and have them read back to you.
 2. [Tutorial Video 2](#): it's not as thorough, but it is a bit easier to see the menu
 3. [Tutorial Video 3](#). Because why not a third one?
2. Text-to-Speech in Google Docs
 - If you prefer working in Google Docs you can do the same thing. [This article](#) will show you how. You can also watch [this tutorial video](#).

Better still: 'speak-on-enter' to confirm data values as entered

1. speak-on-enter [[tutorial:](#)]

Speech-to-Text for keyboardless data entry

1. Google Sheets: [Typing with your voice](#)
2. MS Word: [How to dictate documents](#)

Overview of MS Accessibility Tools

1. 'Narrator' function [link](#)

Readings

1. Campbell, J. L. *et al.* 2013. Quantity is nothing without quality: automated QA/QC for streaming environmental sensor data. *BioScience*, 63(7): 574-585. [link](#)

For more advanced users of R/Github

1. Kim, A. Y. *et al.* 2022. Implementing GitHub Actions continuous integration to reduce error rates in ecological data collection. *Methods in Ecology and Evolution*, 13, 2572– 2585. <https://doi.org/10.1111/2041-210X.13982>

R Packages for Data QA/QC

1. **janitor**: simple functions for examining and cleaning dirty data. It was built with beginning and intermediate R users in mind and is optimized for user-friendliness. Advanced R users can already do everything covered here, but with janitor they can do it faster. (see this [mini-tutorial](#))
 - clean column names
 - remove empty rows and columns
 - remove duplicated rows
2. **cleanr**. small R package for cleaning and checking data columns in a fast and easy way.
3. **unheader**: used to wrangle spreadsheets with embedded subheaders or values wrapped across several rows (highlighting, merged cells, etc).

For more advanced R users:

1. Richard's Iannone **pointblank** R package for Data Validation and Organization of Meta-data [link](#)
2. **validate**: designed to test data against a reusable set of data validation rules, investigate, summarize, and visualize data validation results, among other things
3. **Data Curator** is a simple desktop data editor to help describe, validate and share usable open data.
4. Kim, A. Y., Herrmann, V., Barreto, R., Calkins, B., Gonzalez-Akre, E., Johnson, D. J., Jordan, J. A., Magee, L., McGregor, I. R., Montero, N., Novak, K., Rogers, T., Shue, J., & Anderson-Teixeira, K. J. (2022). Implementing GitHub Actions continuous integration to reduce error rates in ecological data collection. *Methods in Ecology and Evolution*, 13, 2572–2585. <https://doi.org/10.1111/2041-210X.13982>

Sources

DataONE Education Module: Data Quality Control and Assurance. DataONE. Retrieved Nov12, 2012. From http://www.dataone.org/sites/all/documents/L05_DataQualityControlAssurance.pptx

Barchard, K. A., & Pace, L. A. (2011). Preventing human error: The impact of data entry methods on data accuracy and statistical results. *Computers in Human Behavior*, 27(5), 1834-1839. [link](#)

Atkinson, I. (2012). Accuracy of data transfer: double data entry and estimating levels of error. *Journal of Clinical Nursing*, 21(19pt20), 2730-2735. [link](#)

Goldberg, S. I., Niemierko, A., & Turchin, A. (2008). Analysis of data errors in clinical research databases. AMIA Annual Symposium proceedings. AMIA Symposium, 2008, 242–246. [link](#)