Data Management – Architecture

Claire Osgood

November 2017

CEHI
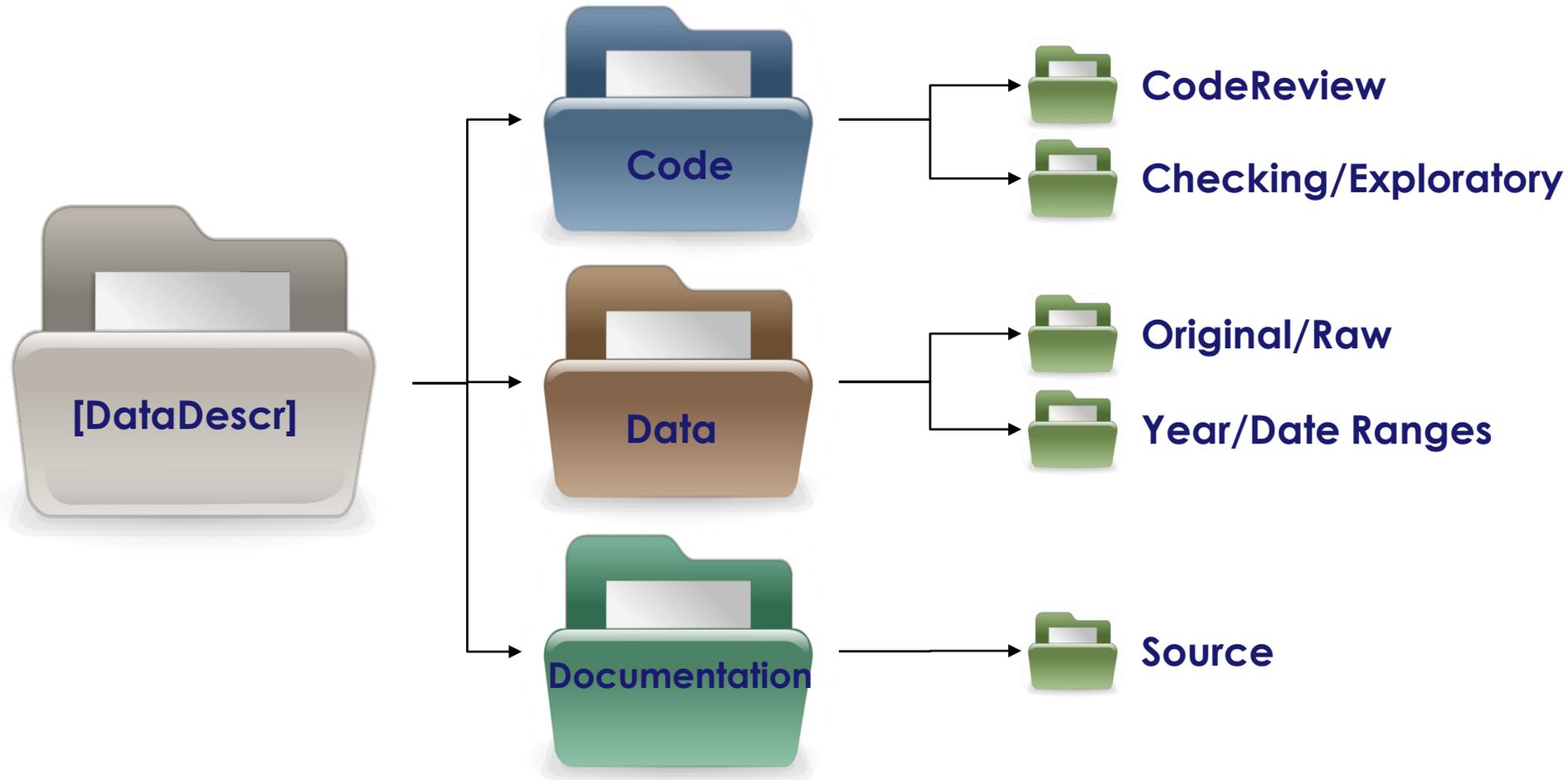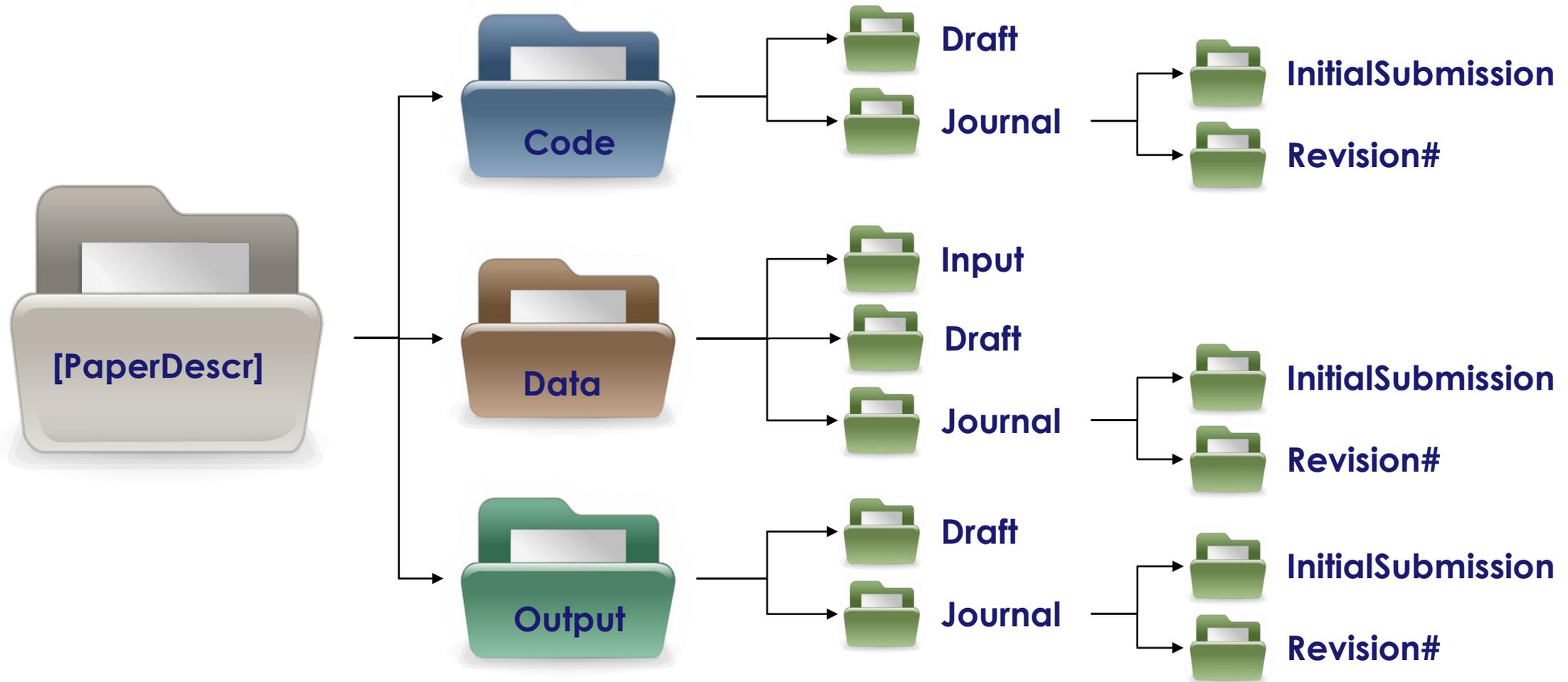Children's Environmental Health Initiative
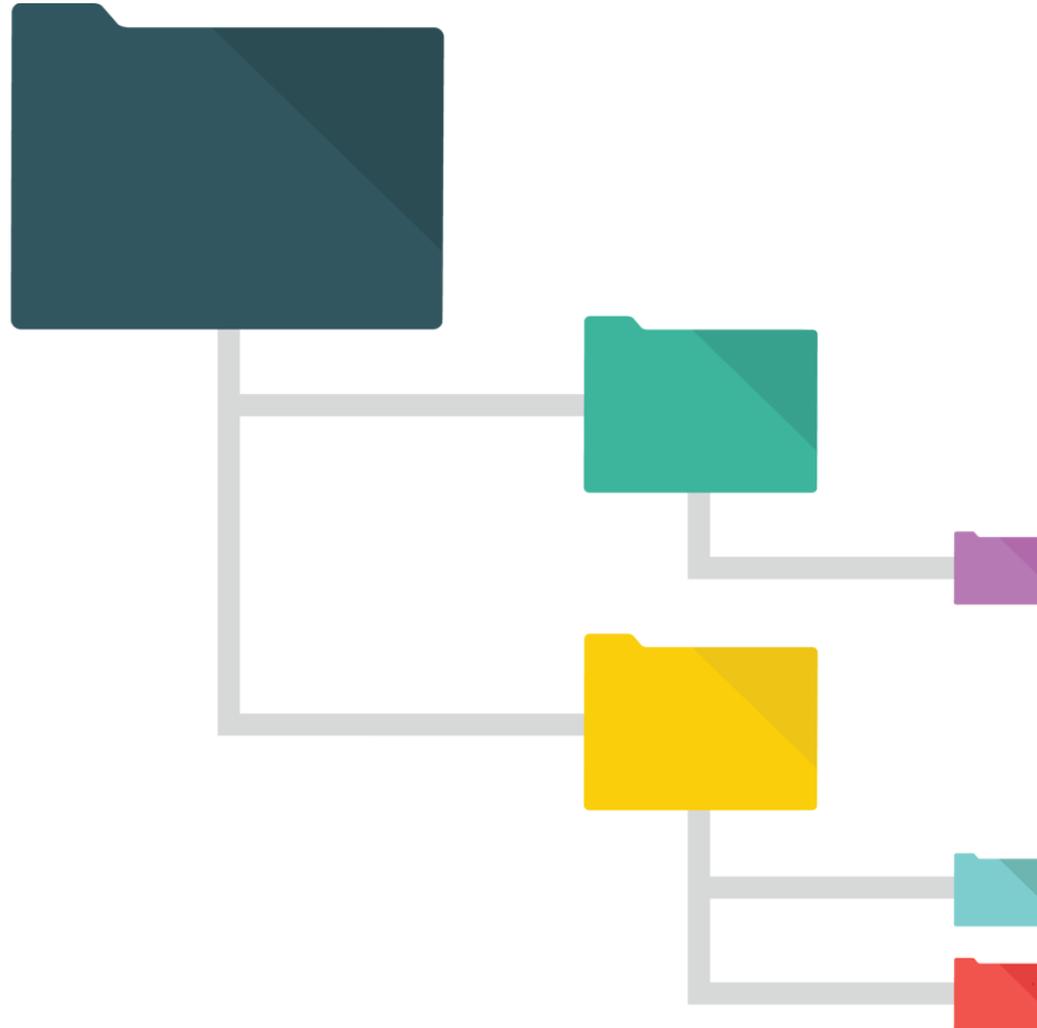
Architecture

Architecture – Folder Structure for Data Prep

Architecture – Folder Structure for Statistical Analysis

From "CEHI_Statistical_Analysis_Guidelines_2016-12-20"

Architecture – File and Folder Naming

DO:
- Use underscores for spaces
- Format dates yyyy-mm-dd or yyyy-mm
- Use leading zeros for dates and numbers
- Use years/months the data cover
- Describe the contents
- Use standard program prefixes:
    - Read*
    - Extr* or X*
    - Chk*
    - Cr*
    - Rq*
- Include the following for data files:
    - Geographic extent (unless part of folder name)
    - Date(s) of coverage
    - Subject/content

Architecture – File and Folder Naming

DON'T:
- Use spaces or special characters
- Format dates mm-dd-yyyy or mm-yyyy
- Use date file was received
- Use personal names
- Use "final", "new", "data", or default names
- Repeat info in the parent folder name
    - Caveat – sometimes it is appropriate to repeat some of the folder information

<u>Architecture – File and Folder Naming – Quiz</u>

Which is better, A or B? Why?

For cardiovascular data covering 2009-2011, received June 2012:

A: Cardio_2009_2011.xlsx   B: For Claire June 2012.xlsx

Architecture – File and Folder Naming – Quiz

Which is better, A or B? Why?

For updated data on Harris County Churches in 2015:

A: Harris_new!data!

B: Harris_Churches_2015

Architecture – File and Folder Naming – Quiz

Which is better, A or B? Why?

For Lead records that had x/y and GIS was used to attach census block:

A: Lead_2014_GeoID

B: Export_Output

Architecture – File and Folder Naming – Quiz

Which is better, A or B? Why?

For notes on questions for the data provider, and their answers:

A: QandA_20160815   B: Q&A_81516

Architecture – File Naming – Collaborative Editing

Guidelines for collaboratively edited files (Word docs):

- First person naming file, or the person designated as the "Keeper of the Document" (KOD), numbers the version (ex: file_1)

- Each subsequent editor of the file makes suggested changes using the track changes options and adds their initials to suffix of the file (ex: file_1_js; file_1_js_kt)

- Once the file has been edited by all members of the edit team, the KOD decides which changes to retain and which to reject and then changes the version number as appropriate. (ex: file_1_js_kt becomes file_2)

Architecture – Additional Resources

For projects including statistical analysis, see additional documents:

CEHI_Statistical_Analysis_Guidelines_2016_12_20.pdf
Includes information on standard folder structure for analysis files and programs.

CEHI_Naming_Conventions_Guidance_2016_06_28.pdf
Reference for file naming conventions.