



# Date Time checking in PASTA

2017 November  
Environmental Data Initiative (EDI)



# Reminders

- 2 Checks - released April 2017
  - Checksum present
  - Checksum matches
- Portal updates - Fall 2017
  - Option to skip data upload if checksum matches
    - Introduction/Tutorial: Nov 28
- Suggestion:
  - Mini-workshop or EDI tutorial
    - Methods to generate and insert a checksum
    - Interpretation

# Outline

1. `DateTimeFormatStringPreferred`
2. `DateTimeCongruence`

# ECC Working group activities, Summer/Fall 2017

Analyzed current practice

Developed 2 checks

formatStringPreferred

- ISO recommended
- Draft: [https://github.com/EDIorg/ECC/blob/master/practices/dateTimeFormatString/best\\_practice.md](https://github.com/EDIorg/ECC/blob/master/practices/dateTimeFormatString/best_practice.md)

formatString-data congruence

- Parsable formats allow more sophisticated data analysis
- E.g, data frequency, gaps, coverage

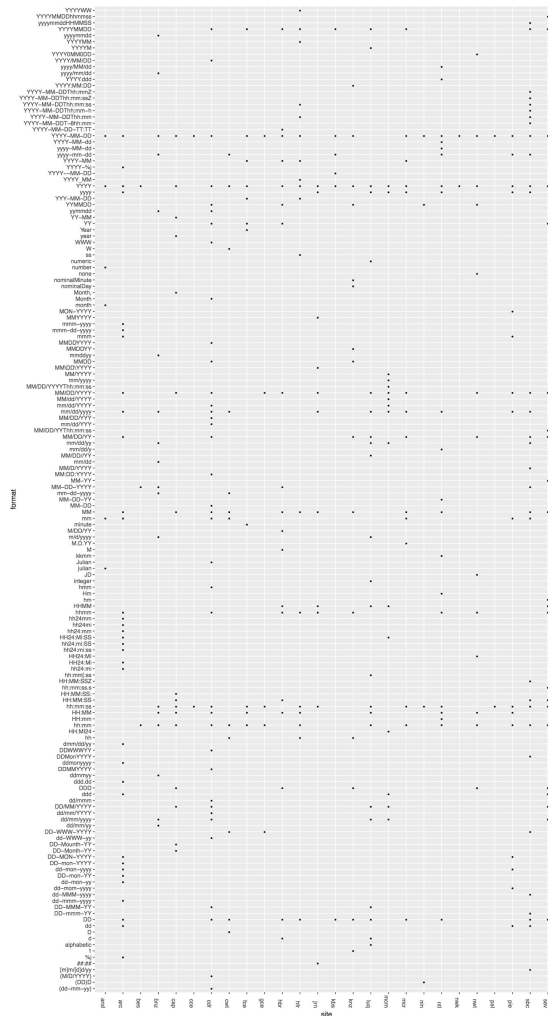
## Current practice:

## At LTER sites

# Total unique dateTime formatStrings: 137\*

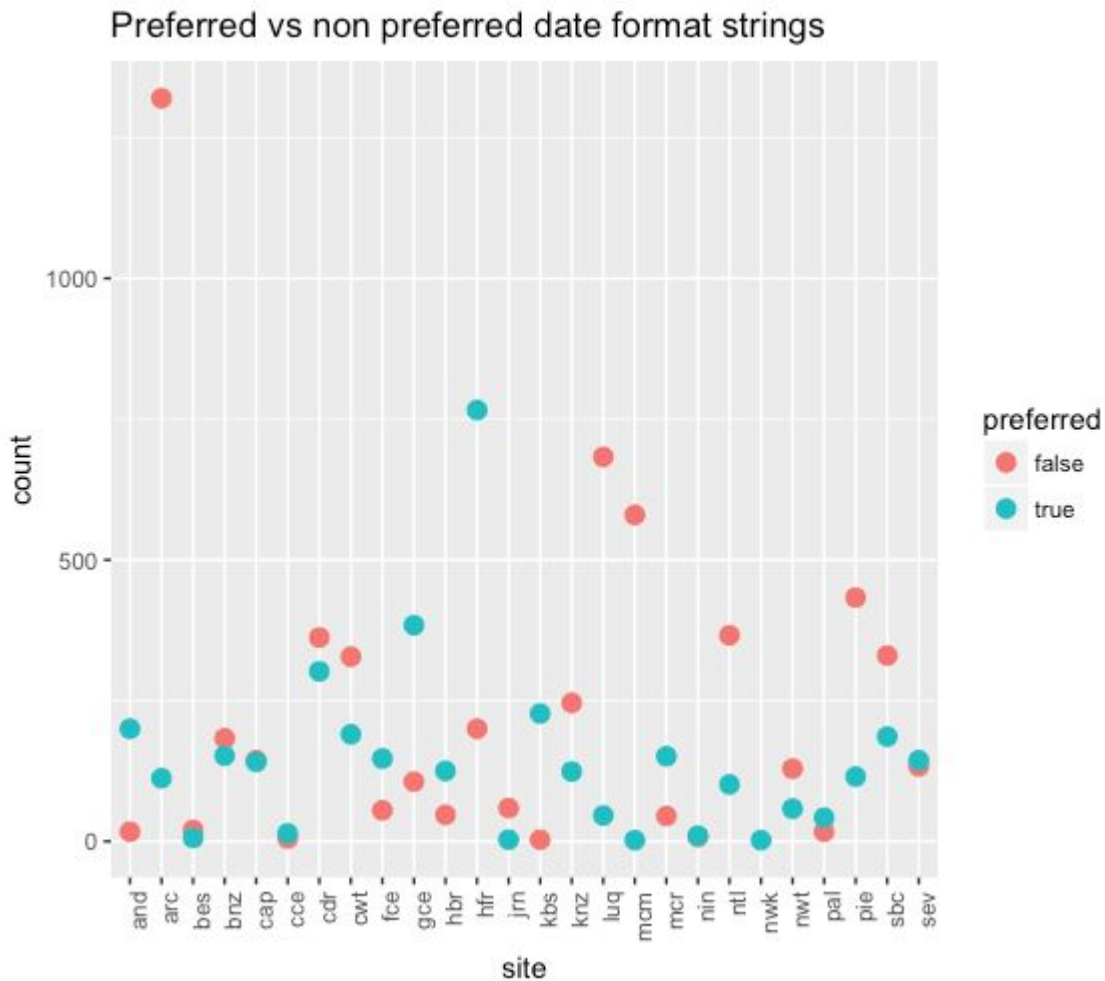
View/download at:

[https://github.com/EDlorg/ECC/blob/master/practices/dateTimeFormatString/DateFormatGraphic1\\_practice\\_sven.pdf](https://github.com/EDlorg/ECC/blob/master/practices/dateTimeFormatString/DateFormatGraphic1_practice_sven.pdf)



## Current Practice:

Among LTER sites, roughly half of `dateTime` `formatStrings` are not on the preferred list



# Check 1: DateTime formatString preferred

Compares content of  
dateTime/formatString element to  
a list constructed from  
combinations of these  
components

Nov 15: *in staging only*

Currently returns “warn”\*

*Behavior and deployment in  
production TBD*

Date	Time	Time zone
YYYY-MM-DD	hh:mm:ss.sss	Z
YYY	hhmmss.sss	+hh:mm
YYYYMMDD	hh:mm:ss	+hhmm
YYYY-MM	hhmmss	+hh
YYYYMM	hh:mm	-hh:mm
YYYY-DDD	hhmm	-hhmm
YYYYDDD	hh	-hh

## Check 2: DateTime Congruence

Compares *any* formatString provided to *every* individual data value

Attempts to parse and confirm correctness

Nov 15: *in staging only*

Returns “warn”

*Deployment in production TBD*



# Recap: Next steps - EDI

- Nov 15: Both checks deployed
  - to “Staging” environment only
  - production on hold
- Develop mechanism to highlight certain “info” messages as “Alert”
- Develop tools to assist converting data values to ISO
  - Possible approaches
    - Before package submittal
    - As a derived dataset
    - Others possible, TBD
  - All approaches require community feedback and engagement
- Suggest additions to EML schema for v 2.2
  - <formatString> system attribute to aid parsing
  - Improved documentation, to encourage ISO

# Recap: Next steps - Data Managers

- Test your datasets at:
  - <https://portal-s.edirepository.org> or
  - <https://portal-s.lternet.edu>
- Comment on recommendation draft:
  - [https://github.com/EDlorg/ECC/blob/master/practices/dateTimeString/best\\_practice.md](https://github.com/EDlorg/ECC/blob/master/practices/dateTimeString/best_practice.md)
  - Discuss on Slack channel [https://edi-got-data.slack.com/#congruence\\_checker](https://edi-got-data.slack.com/#congruence_checker)
- Feedback on implementation for preferred dateTimes:
  - “Warn” or “info/alert”?