CONNECTICUT
Public Health

Electronic Case Reporting Data Validation and Quality Assurance in Connecticut in 2023: Use of PowerShell, XSLT, and Power Query to Improve Efficiency and Timeliness

Mike Schneider, MPH
Sue Speers, MPH
Nancy Barrett, MPH MS
Connecticut Department of Public Health

Disclosures

- No financial interests
- Not an expert
- Not better

What's the problem?

Needs:

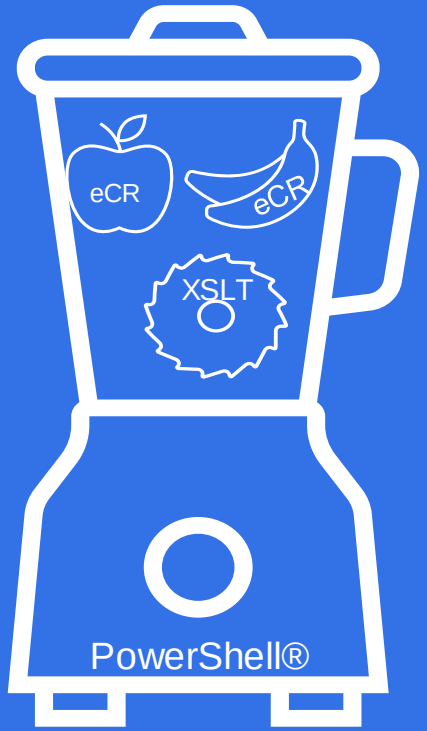
- Testing
- Data validation
- Quality Assurance
- Efficiency

Challenges:

- Small team
- At first, unable to use R
- Later, R scripts were slow

Solution

PowerShell®
XSLT
Power Query®



Power Query®

MS Excel
or CSV file

* XSLT= Extensible Stylesheet Language Transformations

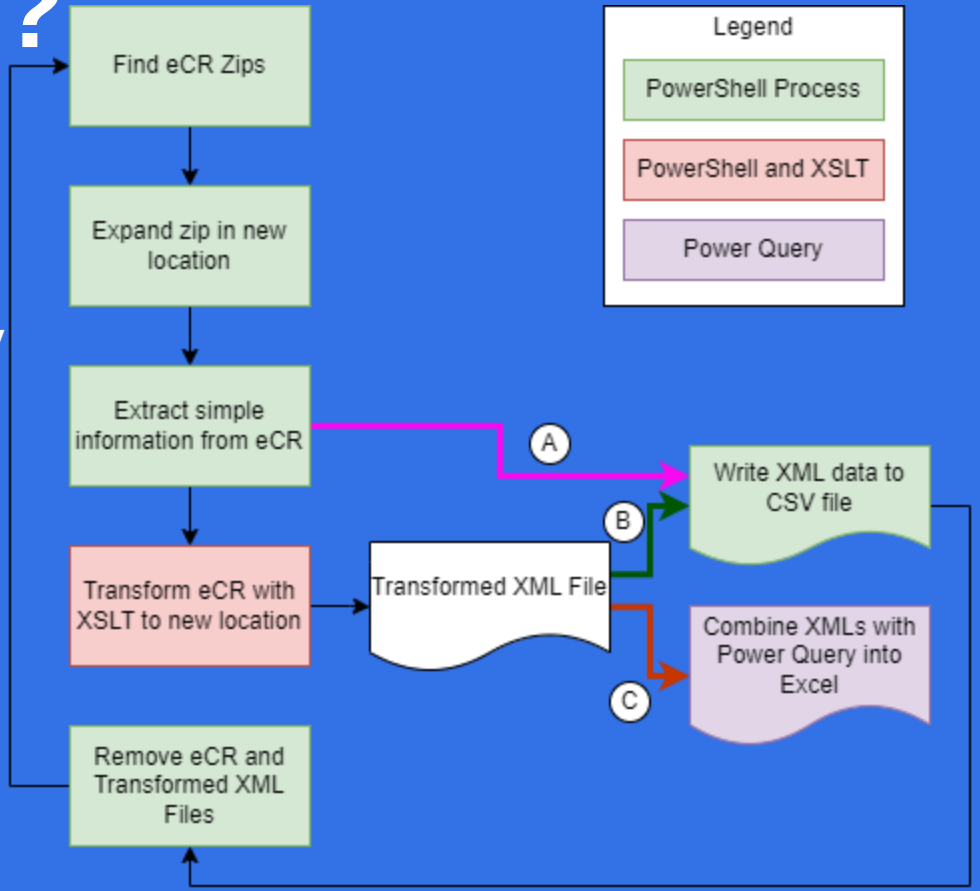
How did we do it?

Three methods used:

A. Pull variables from eCR to CSV
Good for simple, non-repeating elements

B. Transform eCR and Write to CSV
Good for elements that have the potential to repeat and can be made very specific.

C. Transform eCR and use Power Query® to combine output XMLs
Good for **displaying** repeating elements (meds, problems, etc.). Can be very specific if needed.



Data Validation and QA

- It worked!
- Compare elements across different senders
- Sandbox for testing Xpaths
- Able to test rules not in DQ Schematron

HCO	Count of eCRs	Number of Planned Observations	Number of Planned Procedures	Planned Observation per eCR	Planned Procedure per eCR
HCO-A	2,497	23,942	137	9.59	0.05
HCO- B	10,276	0	51,319	0.00	4.99
HCO-C	104	0	863	0.00	8.30
HCO-D	858	4,010	1,864	4.67	2.17
HCO-E	207	0	0	0.00	0.00
Grand Total	13,942	27,952	54,183	2.00	3.89

eCRs received from 01Jan24-01Mar24

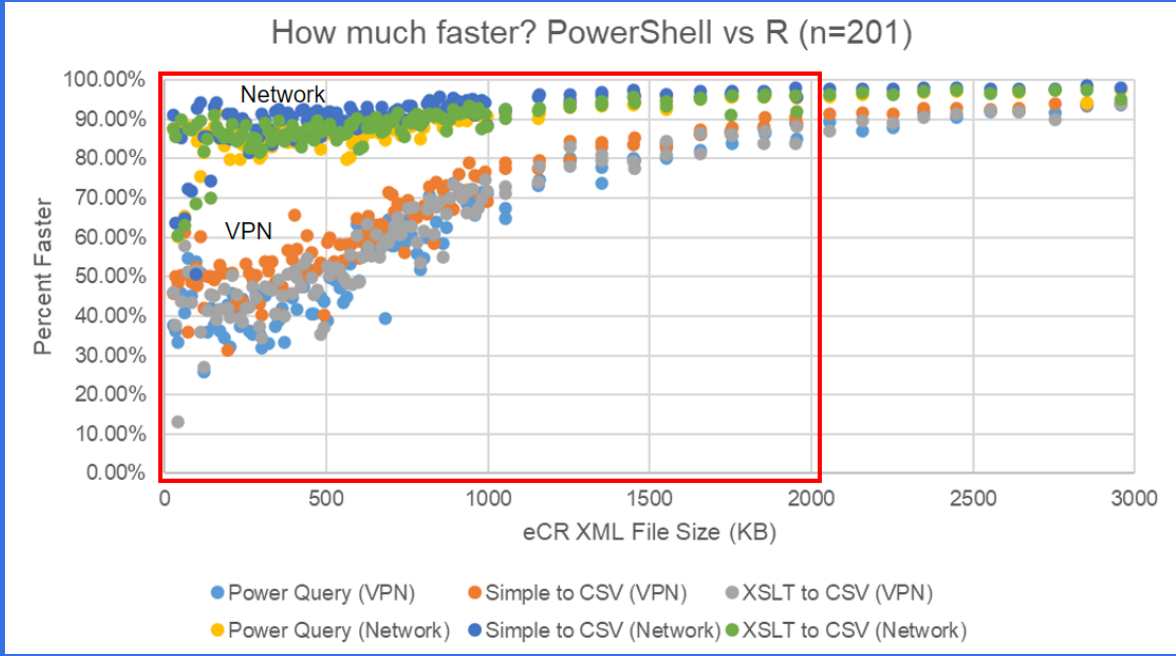
Efficiency

PowerShell® compared to R (parsing the same elements into a CSV file)

- >3X faster over VPN
- >16X faster on the state network

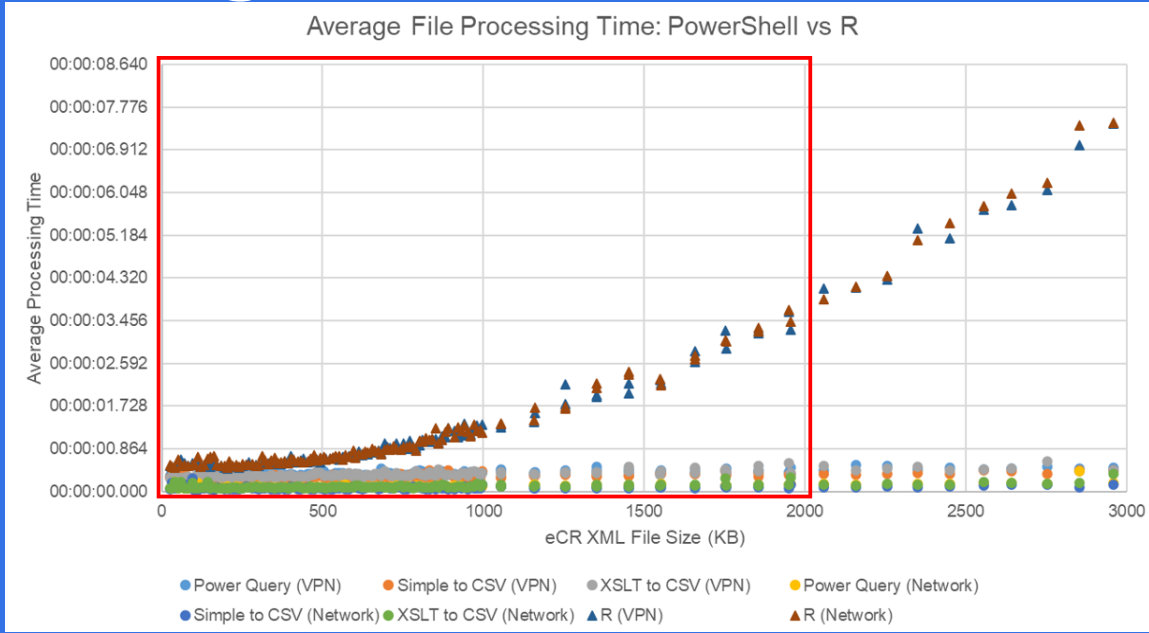
Real Life Example:
In a 15-minute period

- PowerShell® parsed 2,700-15,000 eCRs
- R parsed 900 eCRs



>90% of eCRs in a 3-month period

Processing Time



>90% of eCRs in a 3-month period

Processing Time Ranges, File Sizes 0-2000KB

PowerShell®				R			
VPN		Network		VPN		Network	
Min	Max	Min	Max	Min	Max	Min	Max
00:00:00.256	00:00:00.506	00:00:00.054	00:00:00.247	00:00:00.474	00:00:03.638	00:00:00.480	00:00:03.671

Flexibility

- Search for and copy specific files to a new location
- Transform eCRs into extracted HTMLs for staff
- Search for specific condition codes in RR
- Compare SNOMED codes in eCRs to surveillance system mappings
- Make your script talk

```
Add-Type -AssemblyName System.Speech
$synth = New-Object -TypeName System.Speech.Synthesis.SpeechSynthesizer
$synth.SelectVoice("Microsoft David Desktop")
$synth.Speak("Your script is done, and I am a sentient machine. Bow down to the computer overlords.")
```

Conclusion

- Alternative to R scripts
- Validation/ QA/ Testing
- Efficient and Fast
- The Best Option

Thank you

- Lynn Sosa, MD
- Sue Speers, MPH
- Nancy Barrett, MPH MS
- Informatics Team
- CSTE

Contact and Resources

Mike Schneider
CT Dept of Public Health
Michael.Schneider@ct.gov
DPH.eCRInformatics@ct.gov

https://github.com/mikeeschneider/CSTE2024