

# Automating Open Data Publications in Public Health Scotland

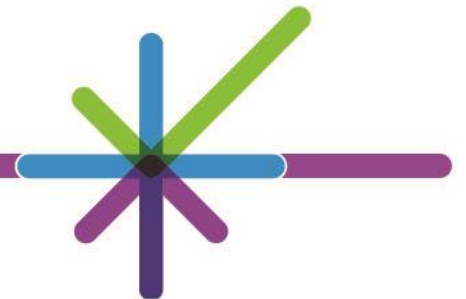
Csilla Scharle

25th November 2021



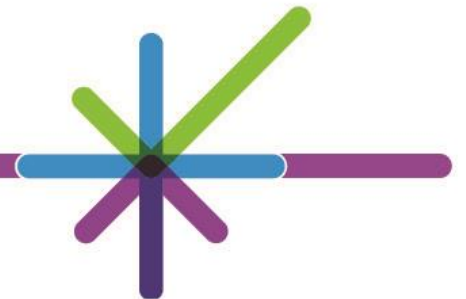
# Open Data platform for PHS

- Scottish Health and Social Care Open Data platform
- CKAN – data management system
  - Open source
  - Python-based
  - Embedded Data API
  - Customisable
  - Rich search functionality
- Growth phase – platform started in 2017
- Currently 90 datasets, 704 resources



# COVID-19

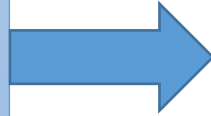
- Demand for timely, reliable and well-structured covid-19 open data
- Rapid development of new datasets:
  - Daily COVID-19 Cases in Scotland
  - COVID-19 Vaccination in Scotland
  - Weekly & monthly datasets
- Increase in volume and frequency
  - 17 files updated daily
  - over 500 files updated monthly



# Consideration for automation

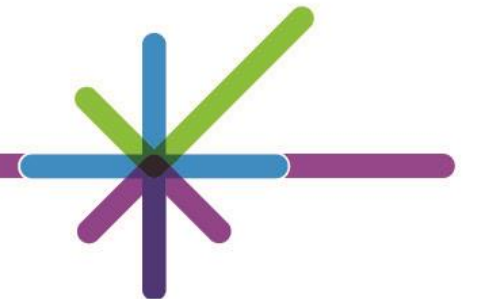
Initial development in python

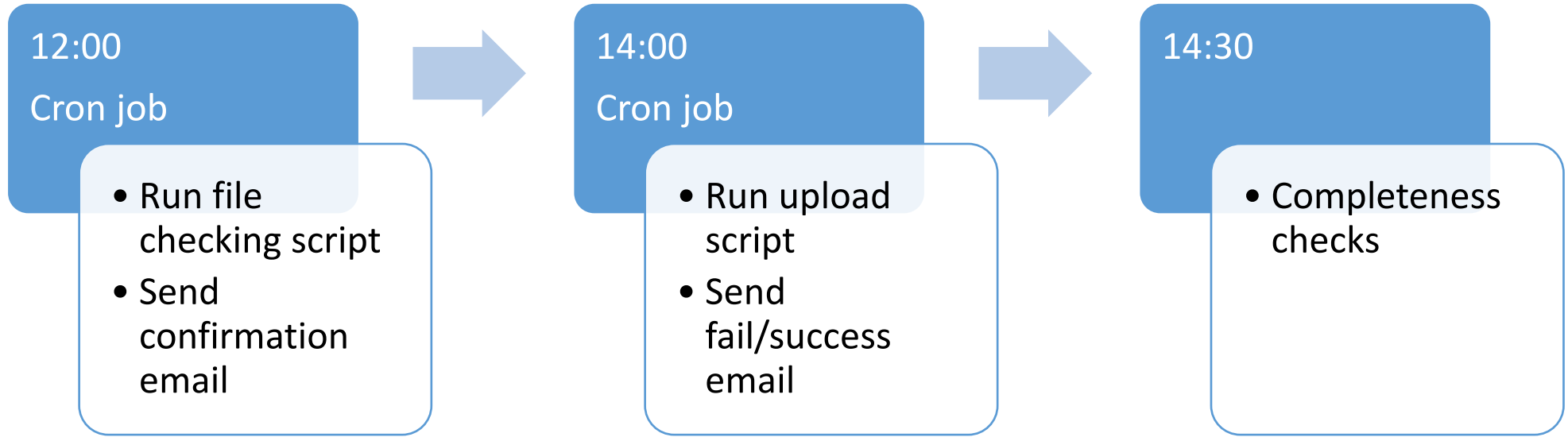
- Compatibility with CKAN
- Efficient batch uploads
- Existing 'ckanapi' package
- Quick development to test API upload



R

- Align with team experience + skillset
- Suits PHS technical infrastructure
- Widely used within PHS – more support available + facilitates future handover
- RStudio Server environment – easy scheduling





# Future Developments

- Robust error-handling & automated retries on failure
- Open-data-maintenance internal R package

File checking  
functions

Diagnostic  
emails

Upload  
functions

Update  
metadata

