Clinical Informatics

How to automate jobs in R and forget about it..

..sort of



July 20

<u>Daria.Cosentino@hcahealthcare.co.uk</u> Clinical.Informatics@hcahealthcare.co.uk

HCAHealthcare uk

Agenda

- 1 Introduction
- 2 Our infrastructure
- 3 A brief history of reporting and its challenges
- 4 A whole "new" world: R and Task Scheduler
- 5 A whole "new" challenge
- 6 ...and they lived happily ever after

Who we are



Director, Clinical Information & Analytics



Clinical Informatics Manager



Senior Clinical **Analyst**



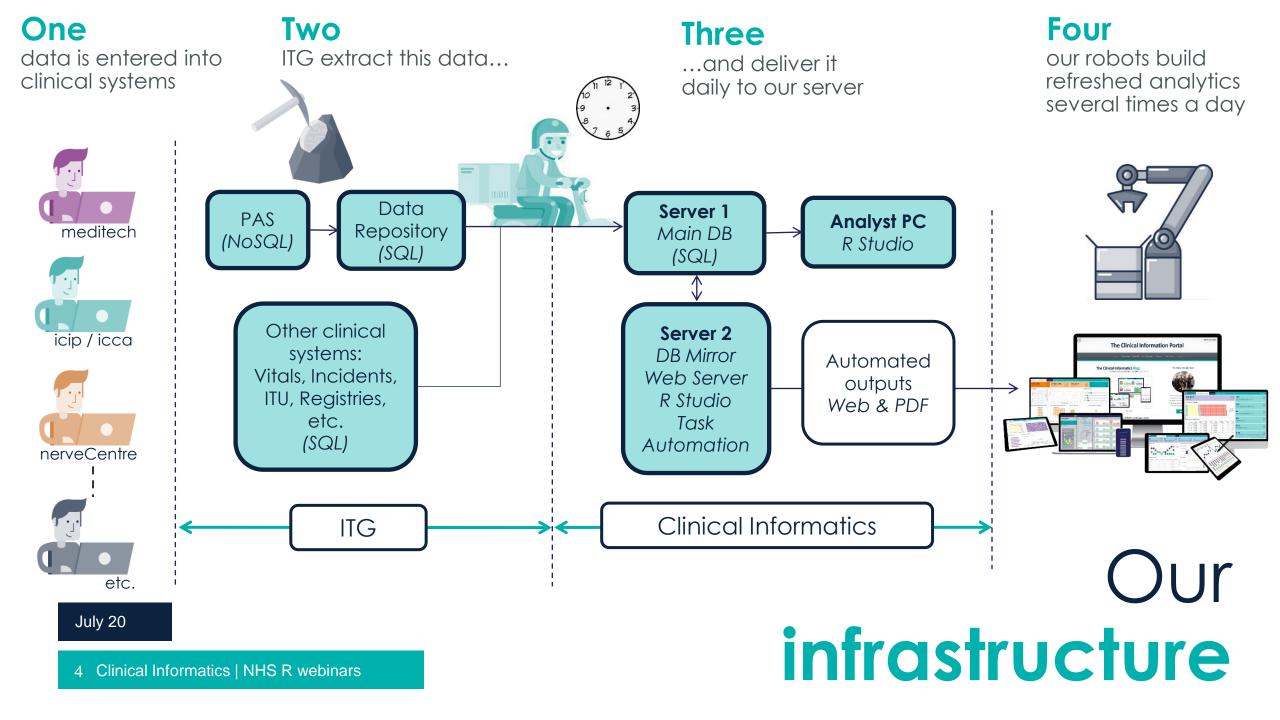


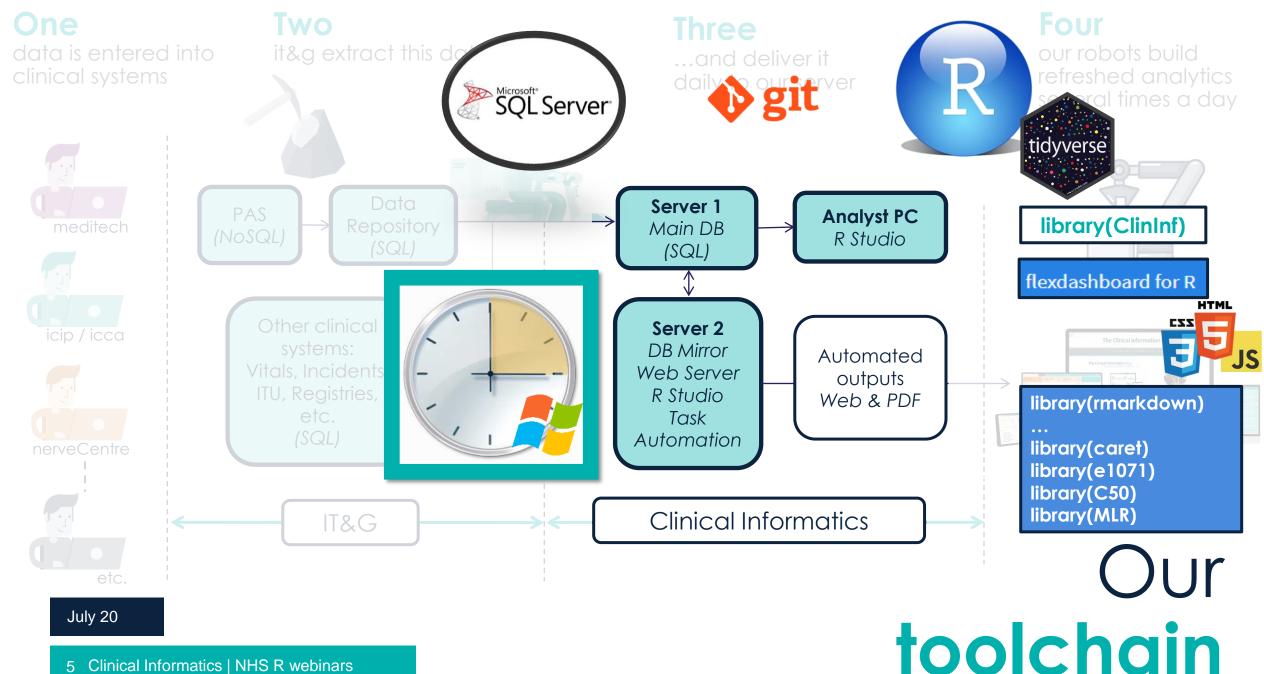




Clinical Analysts & NHS trainee scientist







A brief history of reporting





Excel Connections to SQL



R generated reports



R html reports



CONS

- Prone to errors
- Time consuming
- Frustrating Excel crashing..
- Poor visualisations
- Report can be modified by users
- Data are not live



Less Time consuming

- UX issues
- Updates to the workbook would require sending a new file to the users
- Version Control

- Parametric
- Great versatility
- Version Control
- Great visualisations with ggplot2
- Allows other forms of output

- Distribution still an issue
- Not live as soon as it has been generated (healthcare data is a living organism - end users disputing not correct figures)
- We control the output you can delete the html if it's wrong
- Better UX

The jobs still need to be triggered manually from Rstudio to refresh the output



A whole new world

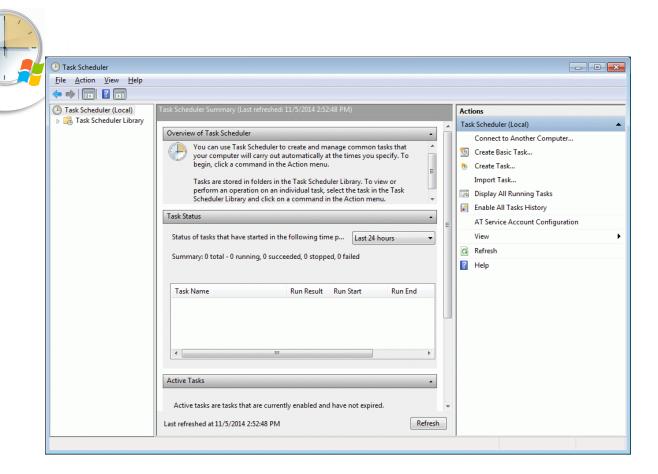
Task Scheduler is a component of Microsoft Windows that provides the ability to schedule the launch of programs or scripts at pre-defined times or after specified time intervals.

It was first introduced for Windows 95 and since then it has been part of any following Windows OS.

The **Start a Program** function within task scheduler work with .exe files, PowerShell or Command Prompt commands.

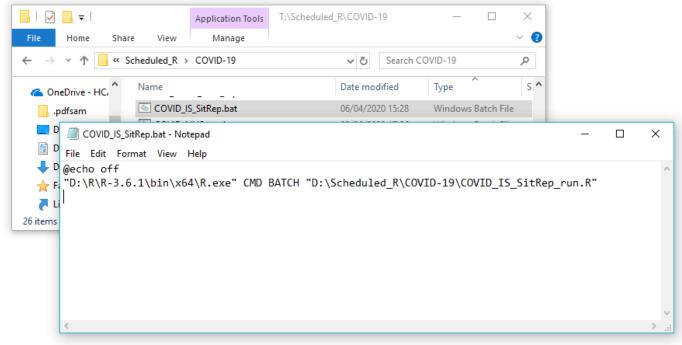
We generally use **Command Prompt** commands saved in a **Batch File** since it is a convenient option to

automate routines.



A Batch File is a special kind of text file that typically has a .bat extension, which can include one or multiple commands that Command Prompt can understand and run in sequence to perform various actions.

A whole new world



To create a **batch** file you only need to:

- 1. Open a **text editor**
- 2. Type the **command** to be run
- 3. Save As .bat

@echo off: disables the display prompt – turns command echoing off

"D:\R\R-3.6.1\bin\x64\R.exe": specifies the Program to be started (just locate on your PC the R.exe file)

CMD BATCH: This batch command invokes another instance of command prompt

"D:\Scheduled_R\....\COVID_IS_SitRep_run.R": path to the R script that will be executed by the program R.exe

July 20

Alternatively, you can use the simplified library(taskscheduleR)

A whole new world

QUICK TROUBLESHOOTING:

1. When using the *library (taskscheduleR)* Add-In make sure to use directories without white spaces or it'll throw an error

Listening on http://127.0.0.1:4838
Warning: Error in observeEventHandler: Make sure the RscriptRepository does not contain spaces, change argument C:/Users/DCosentino/OneDrive - HCAI/Desktop/NHS_R_webinar to another location on your drive which contains no spaces
75: stop

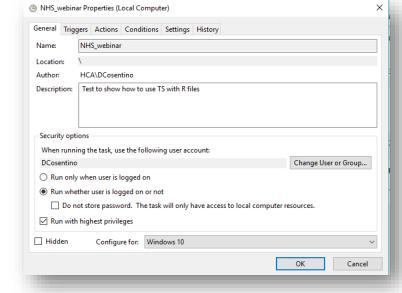
74: observeEventHandler
3: shiny::runApp
2: shiny::runGadget

> taskscheduleR:::taskschedulerAddin()

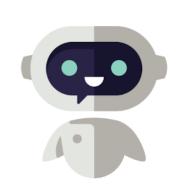
- 2. When using **Task Scheduler + .bat files**, tick the option: "Run whether user is logged on or not" and if you are having folders permission issues (error 0x2), tick "Run with the highest privileges"
- 3. If you want to render a html page (with flexdashboard or rmarkdown) you most likely will have to specify the location of PANDOC (which is called only from the IDE Rstudio, not from R.exe) at the beginning of your Rscript:

Sys.setenv(RSTUDIO_PANDOC="D:\\R\\RStudio\\bin\\Pandoc")





A whole new challenge



R Jobs Status



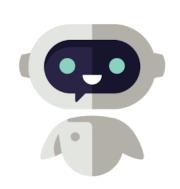
Now that we can automate all the Jobs we want... where is the limit?

The main risk is **JOB Failure**.. And we don't want to receive the news that a job has failed by the end-users themselves!

- Diagnostics Dashboard
- "Jobs Log" Table

08-Jul-2020 Data Diagnostics R Jobs Timeline 07 July 2020 dashWA sql KPIsmaternity sql_dash_cyberknife sql_COVID_NHS_sus dash_blc sal COVID NHS England theatre sql_COVID_NHS_England_beds dash_cc report OCORpacks dash_cc report MCORpacks eport_ClinCodingReports dash_C(notif_datix_escalation notif_COVID_IS_SitRep email_covid_lab_results dashWARD dash train compliance dash theatre dash_KP dash_promsq1 dash_m dash_datadiag dash_PC dash_criticalcare dash CORpages dash ccNHS dash blood HCAHealthcare uk

A whole new challenge

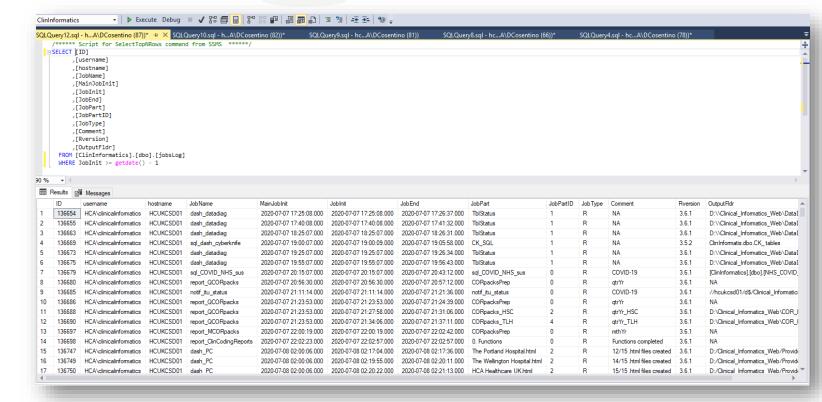




Now that we can automate all the Jobs we want... where is the limit?

The main risk is **JOB Failure**.. And we don't want to receive the news that a job has failed by the end-users themselves!

- Diagnostics Dashboard
- "Jobs Log" Table



Our workflow

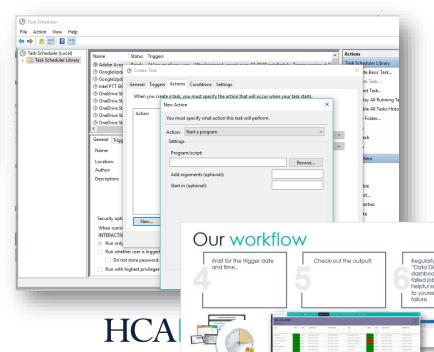
Create the R script you want to be run from Task-Scheduler

July 20

Create the .bat file you want Task-Scheduler to run, specifying the R.exe to use and the path to the Rscript



Create a new Task in Task scheduler, specifying the Action and the Trigger



Our workflow

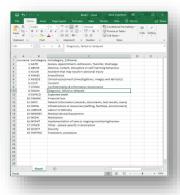
Wait for the trigger date and time..

Check-out the output!

Regularly check the "Data Diagnostics" dashboard to identify failed jobs ... and if helpful send notifications to yourself in case of jobs failure



08-Jul-2020 Next Run Time Last Run Time Next Run Time Last Run Time 08/07/2020 12:05:00 09/07/2020 12:05:00 email_drugalctionar 05/07/2020 04:55:00 12/07/2020 04:55:00 dashWARD CaseMi notif CI train compliant 08/07/2020 09:19:24 04/08/2020 17:00:0 notif_COVID_IS_SitRep 09/07/2020 09:15:0 dash acutemedicine dash_blood 08/07/2020 14:00:0 dash_ccmonitorin 08/07/2020 13:10:00 08/07/2020 14:10:0 12/07/2020 20:00:0 dash_ccNHS 09/07/2020 04:24:00 07/07/2020 14:51:35 dash CORpage report ClinCodingRepo 08/07/2020 22:02:23 08/07/2020 16:00:00 07/07/2020 22:02:22 **zlir** 08/07/2020 22:00:0 dash_criticalcare 09/07/2020 07:45:00 report_MCORpack 07/07/2020 22:00:0 dash_datadiag 10/07/2020 20:00:0 mailR) dash_FAQpedia 07/07/2020 21:00:22 08/07/2020 21:00:0 dash_KPIval 08/07/2020 02:45:00 09/07/2020 02:45:0 dash_maternity 08/07/2020 23:30:00 20/06/2020 21:20:16 12/07/2020 21:20:1 08/07/2020 02:45:00 09/07/2020 02:45:00 30/04/2020 22:30:00 15/07/2020 22:30:0 dash_moor opc dash_medrec sql_bed_bureau_app dash_NPR sql_COVID_NHS_England_bed 08/07/2020 12:20:00 09/07/2020 12:20:00 08/07/2020 18:20:0 dash PC 09/07/2020 02:00:00 sal COVID NHS England theat 07/07/2020 18:20:00 09/07/2020 07:00:00 sal COVID NHS sus 07/07/2020 20:15:00 08/07/2020 20:15:00 dash pharmacy dash_preassessmen 08/07/2020 18:01:51 sql_dash_cyberknife 07/07/2020 19:00:00 08/07/2020 19:00:0



.xlsx .csv
library(writexl)



242 Marylebone Road London, NW1 6JL <u>Daria.Cosentino@hcahealthcare.co.uk</u> <u>Clinical.Informatics@hcahealthcare.co.uk</u> www.hcahealthcare.co.uk

July 20

14 Clinical Informatics | NHS R webinars