

University of  
Sheffield



Slido: #052578

# Introduction

Software Reengineering  
(COM3523 / COM6523)

The University of Sheffield

# Three recent software stories

# CovidSim

A scientific Epidemiological model developed at Imperial College.

<https://github.com/mrc-ide/covid-sim>

Started out as a flu-spread model from ~2005.

Adapted early 2020 to study spread of COVID-19.

Predicted huge death-toll, prompted first lockdowns.

# CovidSim

A scientific Epidemiological model developed at Imperial College.

<https://github.com/mrc-ide/covid-sim>

Started out as a flu-spread model from ~2005.

Adapted early 2020 to study spread of COVID-19.

Predicted huge death-toll, prompted first lockdowns.

Built upon legacy code.

A single ~15,000 LOC undocumented **C**-file.

Lots of machine-translated **Fortran** code.

No systematic tests, only a few sample scenarios.

# CovidSim

A scientific Epidemiological model developed at Imperial College.

Started out as a flu-spread model from ~2005.

Adapted early 2020 to study spread of COVID-19.

Predicted huge death-toll, prompted first lockdowns.

Built upon legacy code.

A single ~15,000 LOC undocumented C-file.

Lots of machine-translated **Fortran** code.

No systematic tests, only a few sample scenarios.

<https://github.com/mrc-ide/covid-sim>



# CovidSim

A scientific Epidemiological model developed at Imperial College.

Started out as a flu-spread model from ~2005.

Adapted early 2020 to study spread of COVID-19.

Predicted huge death-toll, prompted first lockdowns.

Built upon legacy code.

A single ~15,000 LOC undocumented C-file.

Lots of machine-translated **Fortran** code.

No systematic tests, only a few sample scenarios.

Prompted debate about whether the lockdown was justified.

<https://github.com/mrc-ide/covid-sim>

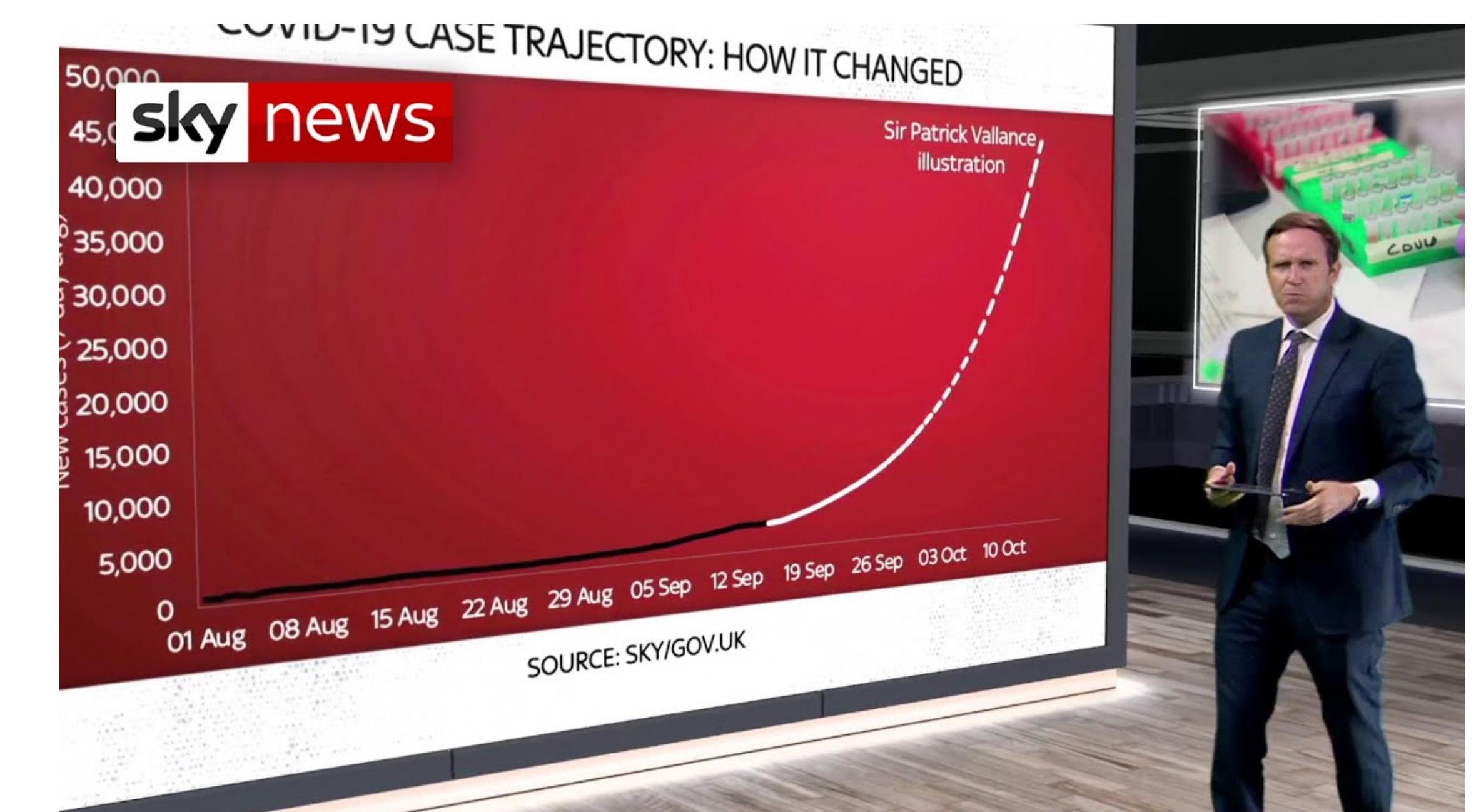


# Test & Trace Data Pipeline Error

PHE relied on an **old Excel format** (XLS) to import the CSV files.

Limited to 65,536 rows.

Patients were missed from CSV files with >65,536 rows.

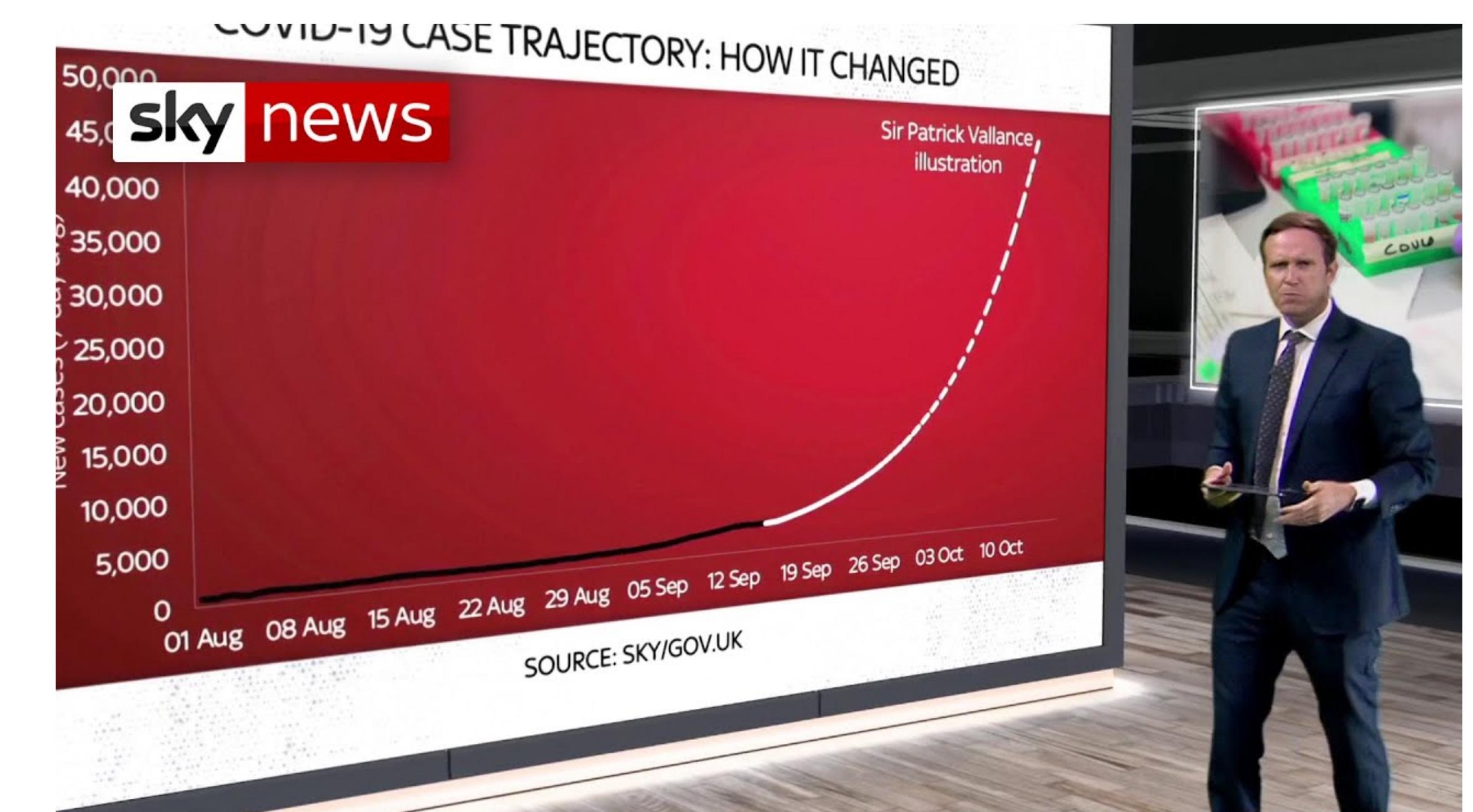


# Test & Trace Data Pipeline Error

PHE relied on an **old Excel format** (XLS) to import the CSV files.

Limited to 65,536 rows.

Patients were missed from CSV files with >65,536 rows.



# Test & Trace Data Pipeline Error

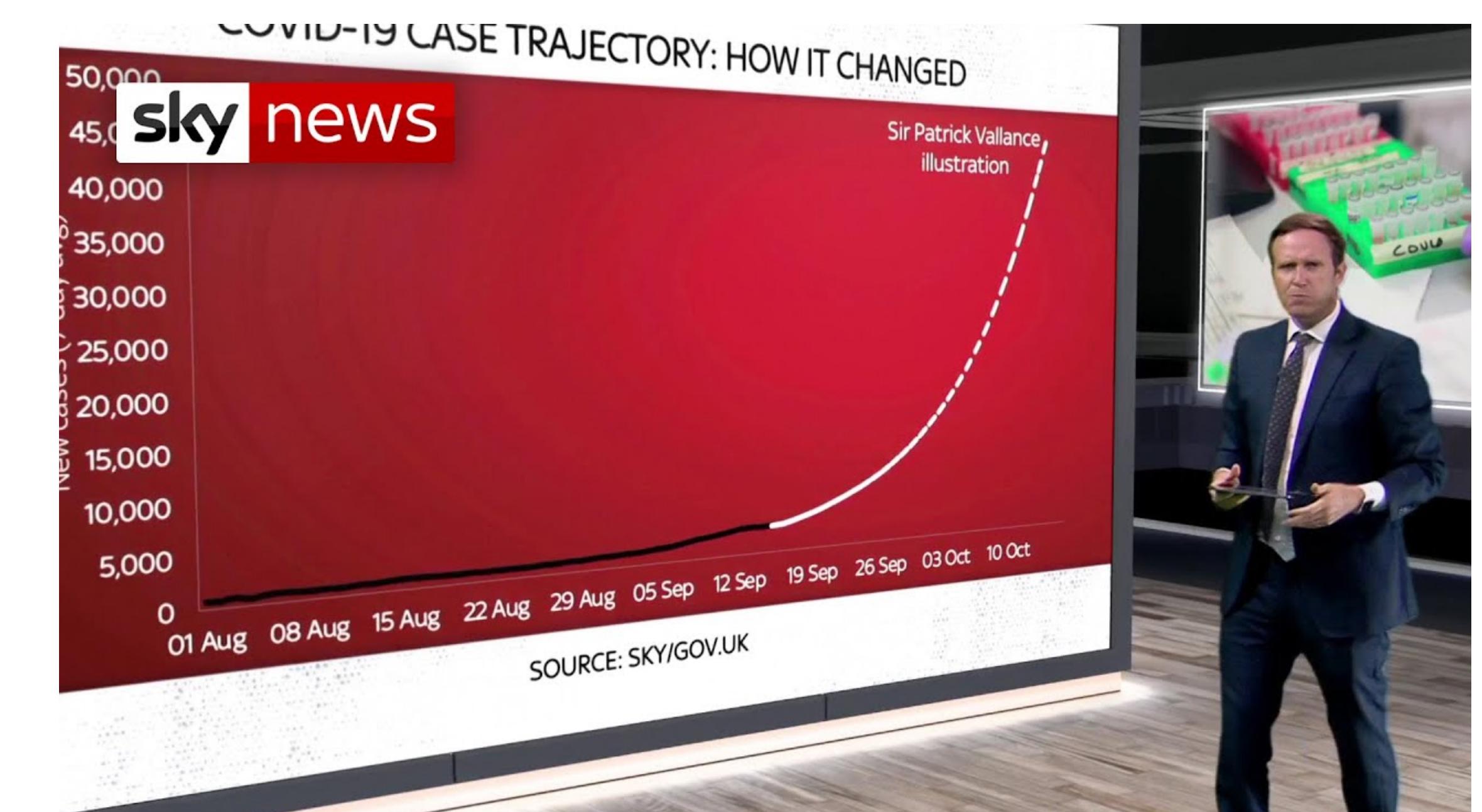
PHE relied on an **old Excel format** (XLS) to import the CSV files.

Limited to 65,536 rows.

Patients were missed from CSV files with >65,536 rows.

Led to an estimated 125,000 additional infections.

Estimated to have caused 1,500 additional deaths.



Fetzer, T., & Graeber, T. (2020). Does contact tracing work? quasi-experimental evidence from an excel error in England.

# Log4J Vulnerability

Vulnerability in highly popular Java logging framework.

Discovered November 2021.

# Log4J Vulnerability

Vulnerability in highly popular Java logging framework.

Discovered November 2021.

Special action taken for messages containing:  
“\${jndi:ldap://LDAP-SERVER/a}”

Would look up log server and execute class at URL.

# Log4J Vulnerability

Vulnerability in highly popular Java logging framework.

Discovered November 2021.

Special action taken for messages containing:  
“\${jndi:ldap://LDAP-SERVER/a}”

Would look up log server and execute class at URL.

Fix has been released in an upgrade of library. But...

Dependent applications may take years to migrate to the fixed version.

Migration can incur its own risks, require follow-on upgrades.

# Log4J Vulnerability

Vulnerability in highly popular Java logging framework.

Discovered November 2021.

Special action taken for messages containing:  
“\${jndi:ldap://LDAP-SERVER/a}”

Would look up log server and execute class at URL.

Fix has been released in an upgrade of library. But...

Dependent applications may take years to migrate to the fixed version.

Migration can incur its own risks, require follow-on upgrades.

The screenshot shows a news article from The Independent. The header 'NEWS' is in white on a red background. Below it is a navigation bar with links: Home, Coronavirus, Climate, UK, World, Business, Politics, Tech, Science, Health, and Family & Education. The 'Tech' link is underlined. A red horizontal bar below the navigation bar contains the word 'Technology'. The main title of the article is 'Flaw prompts 100 hack attacks a minute, security company says'. Below the title is a small text '© 13 December'.

# Log4J Vulnerability

Vulnerability in highly popular Java logging framework.

Discovered November 2021.

Special action taken for messages containing:  
“\${jndi:ldap://LDAP-SERVER/a}”

Would look up log server and execute class at URL.

Fix has been released in an upgrade of library. But...

Dependent applications may take years to migrate to the fixed version.

Migration can incur its own risks, require follow-on upgrades.

A screenshot of a CNN Business news article. The header reads "NEWS" and "Technology". The main title is "Flaw prompts 100 hack attacks a minute, security company says". Below the title is a timestamp "13 December".

A screenshot of a CNN Business news article. The header includes the CNN Business logo and navigation links for Markets, Tech, Media, Success, Perspectives, and Videos. The main title is "The Log4j security flaw could impact the entire internet. Here's what you should know". Below the title is the author's name "By Jennifer Korn" and the update time "Updated 1433 GMT (2233 HKT) December 16, 2021".

# Log4J Vulnerability

Vulnerability in highly popular Java logging framework.

Discovered November 2021.

Special action taken for messages containing:  
“\${jndi:ldap://LDAP-SERVER/a}”

Would look up log server and execute class at URL.

Fix has been released in an upgrade of library. But...

Dependent applications may take years to migrate to the fixed version.

Migration can incur its own risks, require follow-on upgrades.



A screenshot of a CNN News article titled "Flaw prompts 100 hack attacks a minute, security company says". The article is dated December 13, 2021. The headline is prominently displayed in large, bold, dark text. Below the headline, there is a sub-headline and some smaller text. The CNN logo is visible in the top left corner of the article area.



A screenshot of a CNN Business article titled "The Log4j security flaw could impact the entire internet. Here's what you should know". The article is by Jennifer Korn and was updated on December 16, 2021. The headline is in large, bold, dark text. Below the headline, there is a sub-headline and some smaller text. The CNN Business logo is visible in the top left corner of the article area.

The ‘most serious’ security breach ever is unfolding right now. Here’s what you need to know.

Much of the Internet, from Amazon’s cloud to connected TVs, is riddled with the log4j vulnerability, and has been for years

# Log4J Vulnerability

Vulnerability in highly popular Java logging framework.

Discovered November 2021.

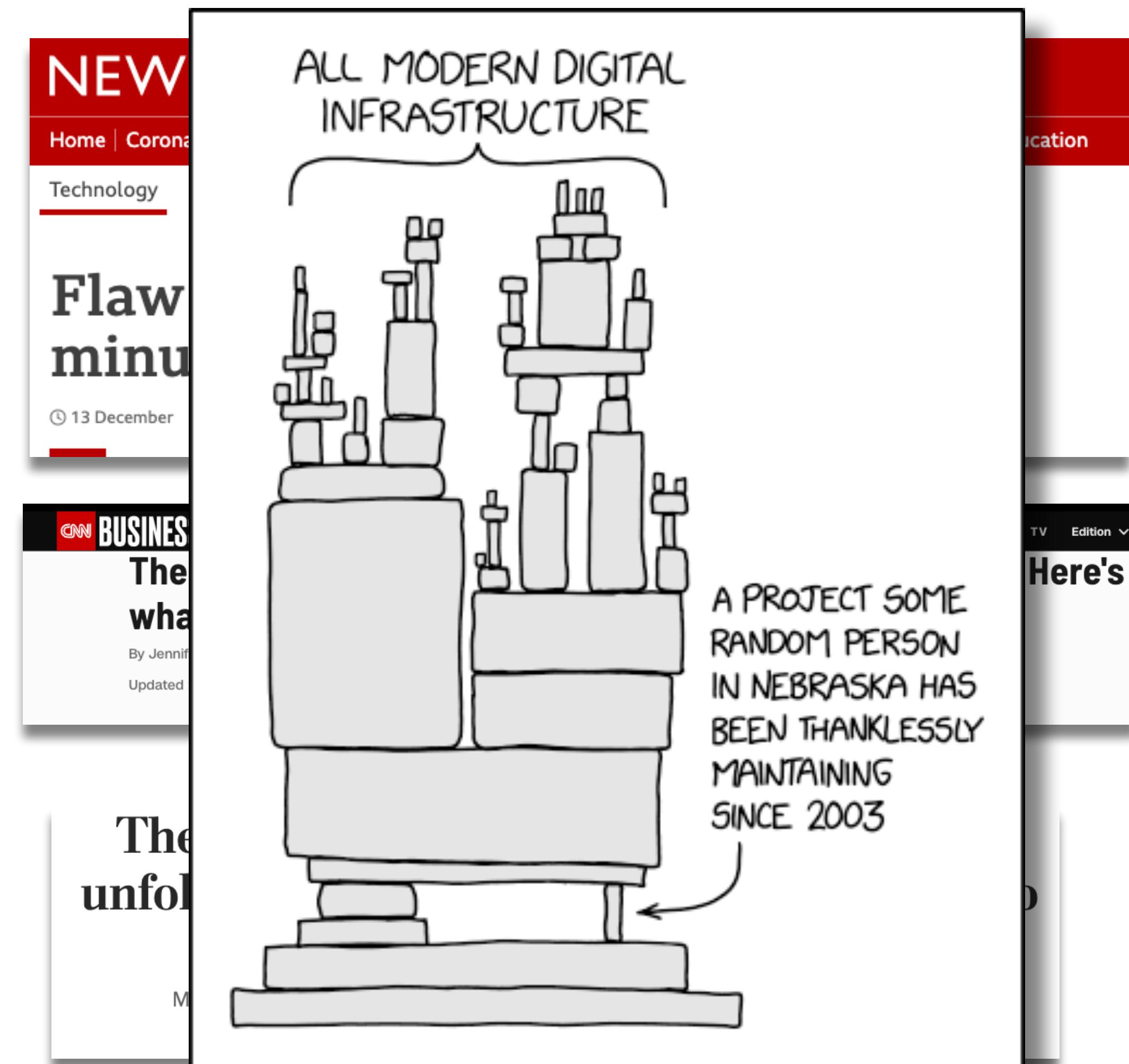
Special action taken for messages containing:  
“\${jndi:ldap://LDAP-SERVER/a}”

Would look up log server and execute class at URL.

Fix has been released in an upgrade of library. But...

Dependent applications may take years to migrate to the fixed version.

Migration can incur its own risks, require follow-on upgrades.



# What's the link?

# What's the link?

They all involve outdated (legacy) software technology.

# What's the link?

They all involve outdated (legacy) software technology.

All have had huge socio-economical impact.

Prompting national lockdowns.

Spreadsheet workflow error led to thousands of preventable infections and deaths.

Huge losses of citizen data through early exploitation of Log4J bug.

# What's the link?

They all involve outdated (legacy) software technology.

All have had huge socio-economical impact.

Prompting national lockdowns.

Spreadsheet workflow error led to thousands of preventable infections and deaths.

Huge losses of citizen data through early exploitation of Log4J bug.

All required an urgent (re-)engineering effort.

CovidSim had not been “renovated” over their lifetimes.

... and the PHE Excel workflow built in legacy technology from the outset.

# What's the link?

They all involve outdated (legacy) software technology.

All have had huge socio-economical impact.

Prompting national lockdowns.

Spreadsheet workflow error led to thousands of preventable infections and deaths.

Huge losses of citizen data through early exploitation of Log4J bug.

All required an urgent (re-)engineering effort.

CovidSim had not been “renovated” over their lifetimes.

... and the PHE Excel workflow built in legacy technology from the outset.

Successful software systems age, and last for decades.

# What's the link?

They all involve outdated (legacy) software technology.

All have had huge socio-economical impact.

Prompting national lockdowns.

Spreadsheet workflow error led to thousands of preventable infections and deaths.

Huge losses of citizen data through early exploitation of Log4J bug.

All required an urgent (re-)engineering effort.

CovidSim had not been “renovated” over their lifetimes.

... and the PHE Excel workflow built in legacy technology from the outset.

Successful software systems age, and last for decades.

**The ability to maintain and renew *big old systems* is an important skill.**

# **Course objectives, delivery and assessment**

# Learning Objectives

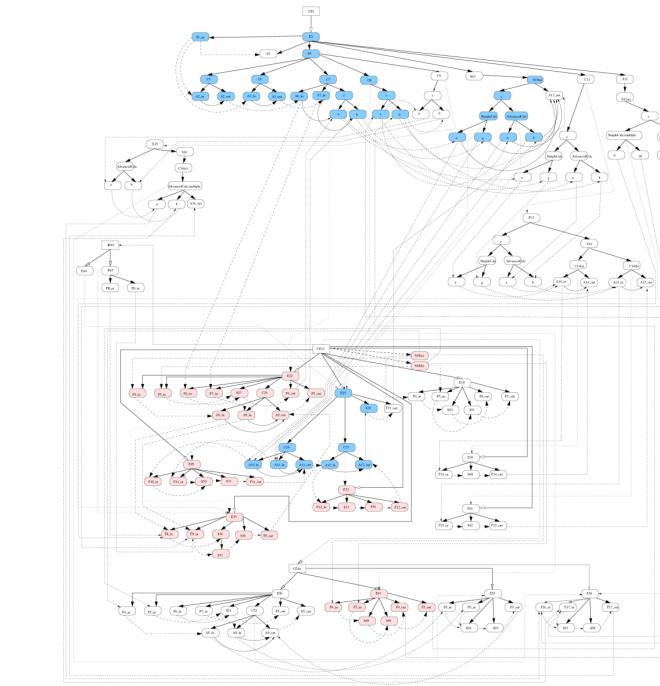
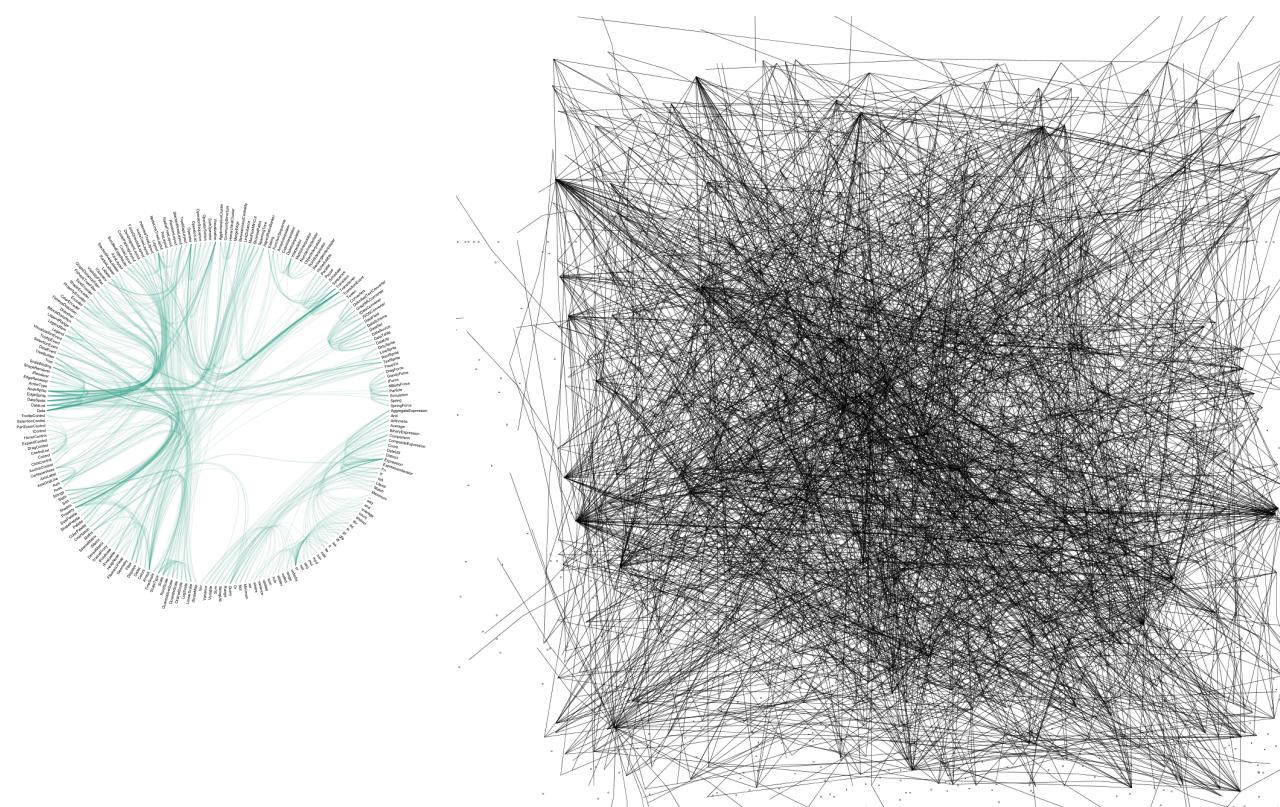
To appreciate the **intrinsic value of legacy systems**.

To **become a considerate coder** - you are writing tomorrow's legacy code.

To learn how to **understand and assess an unfamiliar, complex system**.

To learn about **re-engineering strategies and objectives**.

To **develop your own code analysis tools** to help you along the way.



# Course Delivery

## In-person lectures.

Please bring a device (laptop / mobile phone) so that you can participate in quizzes.

# Course Delivery

## In-person lectures.

Please bring a device (laptop / mobile phone) so that you can participate in quizzes.

## Practical sessions.

Hands-on sessions conducted in a lab.

Often in groups.

# Practical Sessions

## **Weekly lab sessions**

This course is entirely coursework-based.

Participation in lab sessions is mandatory.

Attendance will be taken, and will factor into group-work mark.

# Practical Sessions

## **Weekly lab sessions**

This course is entirely coursework-based.

Participation in lab sessions is mandatory.

Attendance will be taken, and will factor into group-work mark.

## **Group work**

Most of the labs will be carried out within your group.

Use it to catch-up with each other, ask questions,

Your second assignment will be a group-project.

Intended as a “support network”.

# Where can I get support?

# Where can I get support?

## **Sli.do**

During lectures and some lab sessions, we'll run a sli.do Q&A where you can pose your questions and up-vote them.

# Where can I get support?

## **Sli.do**

During lectures and some lab sessions, we'll run a sli.do Q&A where you can pose your questions and up-vote them.

## **Your group**

You will be allocated to a group for the duration of the project.

Intended as a support-group - ask them first!

# Where can I get support?

## **Sli.do**

During lectures and some lab sessions, we'll run a sli.do Q&A where you can pose your questions and up-vote them.

## **Your group**

You will be allocated to a group for the duration of the project.

Intended as a support-group - ask them first!

## **Blackboard discussion forum**

Feel free to post questions here. Will be answered by myself or GTAs.

Please feel free to respond to each others' questions!

GTAs and module leaders will respond in working hours (not over the weekend).

# Where can I get support?

## **Sli.do**

During lectures and some lab sessions, we'll run a sli.do Q&A where you can pose your questions and up-vote them.

## **Your group**

You will be allocated to a group for the duration of the project.

Intended as a support-group - ask them first!

## **Blackboard discussion forum**

Feel free to post questions here. Will be answered by myself or GTAs.

Please feel free to respond to each others' questions!

GTAs and module leaders will respond in working hours (not over the weekend).

**Please only send us emails directly as a last resort.**

# Note on anonymity

Currently, Sli.do and blackboard are anonymous and are not moderated.

Ability for immediate interactions during lectures.

Fast feedback, on-the-spot question up-voting and answering.

This depends on trust - use these fora responsibly.

If this doesn't work, these interactions will have to be moderated.

Would slow down interaction, perhaps remove use of Q&A during lectures.



# Platforms and technology

# Platforms and technology

## Blackboard

Lecture videos and tutorials  
Practical sessions on Blackboard Collaborate  
Discussion fora  
Announcements

# Platforms and technology

**Blackboard**

Lecture videos and tutorials  
Practical sessions on Blackboard Collaborate  
Discussion fora  
Announcements

**GitLab**

Sharing code / solutions.  
Submitting assignments.

# Platforms and technology

**Blackboard**

Lecture videos and tutorials  
Practical sessions on Blackboard Collaborate  
Discussion fora  
Announcements

**GitLab**

Sharing code / solutions.  
Submitting assignments.

**Lab  
technology**

Bash  
Java  
AspectJ  
GraphViz  
Maven  
Git

# Reading material

# Reading material

Lectures will include links to relevant papers or chapters in books.

# Reading material

Lectures will include links to relevant papers or chapters in books.

Three books that cover most of the material in this module:

“*Software Systems Architecture*” by Nick Rozanski and Eoin Woods

A useful reference, but the essential material will be covered in the lectures.

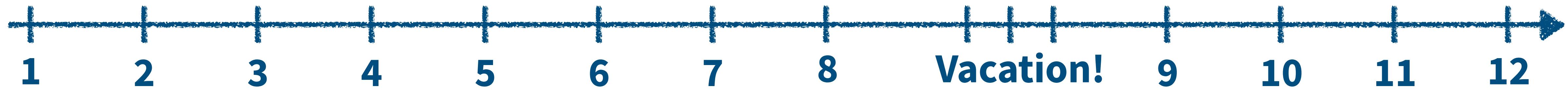
Available online through the library.

“*Object Oriented Reengineering Patterns*” by Serge Demeyer, Stéphane Ducasse, and Oscar Nierstratz.

The book is available as a free e-book here: <http://scg.unibe.ch/download/oorp/>

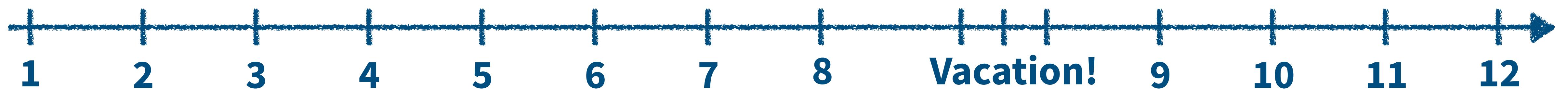
“*Working Effectively with Legacy Code*” by Michael Feathers

Available online through the library.



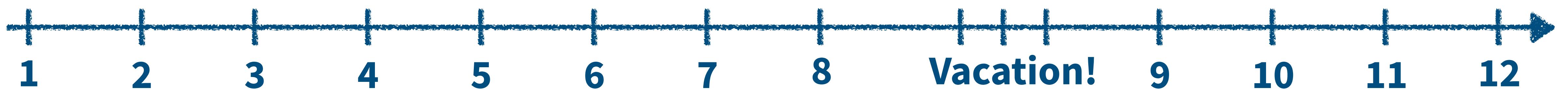
**Intro**

**Legacy  
systems**



**Intro**

**Legacy  
systems**



**Analysing  
source code**

**Intro**

**Legacy  
systems**

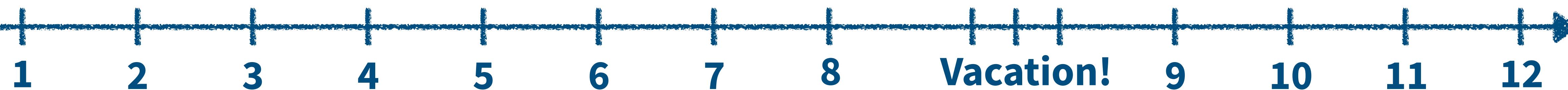
**Analysing  
execution  
data**



**Intro**

**Legacy  
systems**

**Analysing  
execution  
data**



**Intro**

**Legacy  
systems**

**Analysing  
execution  
data**

**Critiquing  
a system**



**Analysing  
source code**

**Version  
repository  
analysis**

**Vacation!**

**Intro**

**Legacy  
systems**

**Analysing  
execution  
data**

**Critiquing  
a system**

**1**

**2**

**3**

**4**

**5**

**6**

**7**

**8**

**Vacation!**

**9**

**10**

**11**

**12**

**Analysing  
source code**

**Version  
repository  
analysis**

**Continuous  
development**

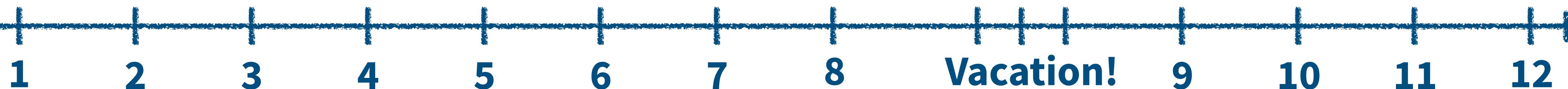
**Intro**

**Legacy  
systems**

**Analysing  
execution  
data**

**Critiquing  
a system**

**Regression  
testing**



**Analysing  
source code**

**Version  
repository  
analysis**

**Continuous  
development**

**Intro**

**Legacy  
systems**

**Analysing  
execution  
data**

**Critiquing  
a system**

**Regression  
testing**

**1**

**2**

**3**

**4**

**5**

**6**

**7**

**8**

**Vacation!**

**9**

**10**

**11**

**12**

**Analysing  
source code**

**Version  
repository  
analysis**

**Continuous  
development**

**Redistributing  
responsibilities**

**Intro**

**Legacy  
systems**

**Analysing  
execution  
data**

**Critiquing  
a system**

**Regression  
testing**

**1**

**2**

**3**

**4**

**5**

**6**

**7**

**8**

**Vacation!**

**9**

**10**

**11**

**12**

**Analysing  
source code**

**Version  
repository  
analysis**

**Continuous  
development**

**Redistributing  
responsibilities**

**Detecting  
duplicate  
code**

**Intro**

**Legacy  
systems**

**Analysing  
execution  
data**

**Critiquing  
a system**

**Regression  
testing**

**Removing  
duplicate  
code**



**Intro**

**Legacy  
systems**

**Analysing  
execution  
data**

**Critiquing  
a system**

**Regression  
testing**

**Removing  
duplicate  
code**

**Reading  
week**



**Analysing  
source code**

**Version  
repository  
analysis**

**Continuous  
development**

**Redistributing  
responsibilities**

**Detecting  
duplicate  
code**

**Reading  
week**

**Intro**

**Legacy systems**

**Analysing execution data**

**Critiquing a system**

**Regression testing**

**Removing duplicate code**

**Reading week**



**Analysing source code**

**Version repository analysis**

**Continuous development**

**Redistributing responsibilities**

**Detecting duplicate code**

**Reading week**

**Assignment 1**

## Intro

## Legacy systems

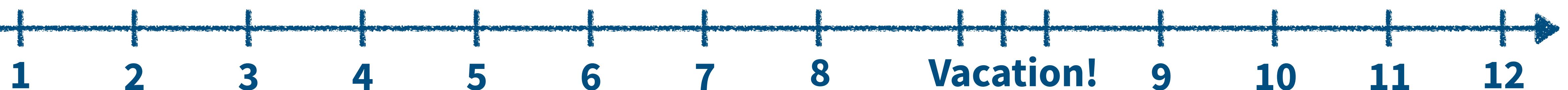
## Analysing execution data

## Critiquing a system

## Regression testing

## Removing duplicate code

## Reading week



Analysing source code

Version repository analysis

Continuous development

Redistributing responsibilities

Detecting duplicate code

Reading week

Assignment 1

Weekly group work

Assignment 2  
(Group Project)

**Intro**

**Legacy systems**

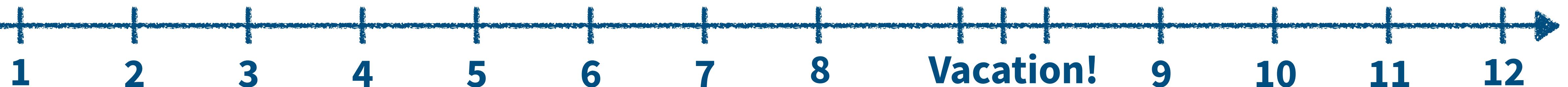
**Analysing execution data**

**Critiquing a system**

**Regression testing**

**Removing duplicate code**

**Reading week**



**Analysing source code**

**Version repository analysis**

**Continuous development**

**Redistributing responsibilities**

**Detecting duplicate code**

**Reading week**

**Assignment 1**



**BB Quiz 1**

**Weekly group work**

**Assignment 2  
(Group Project)**

## Intro

## Legacy systems

## Analysing execution data

## Critiquing a system

## Regression testing

## Removing duplicate code

## Reading week



## Analysing source code

## Version repository analysis

## Continuous development

## Redistributing responsibilities

## Detecting duplicate code

## Reading week

**Assignment 1**



**BB Quiz 1**



**BB Quiz 2**

**Weekly group work**

**Assignment 2  
(Group Project)**



Slido: #052578

# Any questions?

**You can use the Sli.do Q&A if you prefer!**