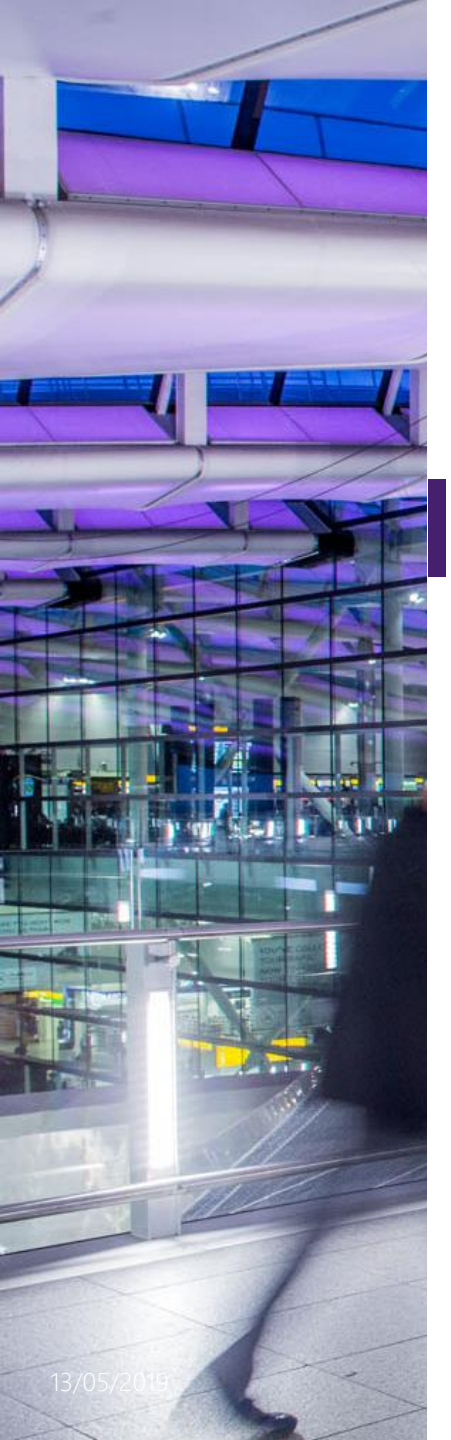
Ilektra Skepetari

Heathrow Forecasting Using R



Agenda

Introduction to Heathrow Forecasting

Forecast Requirements

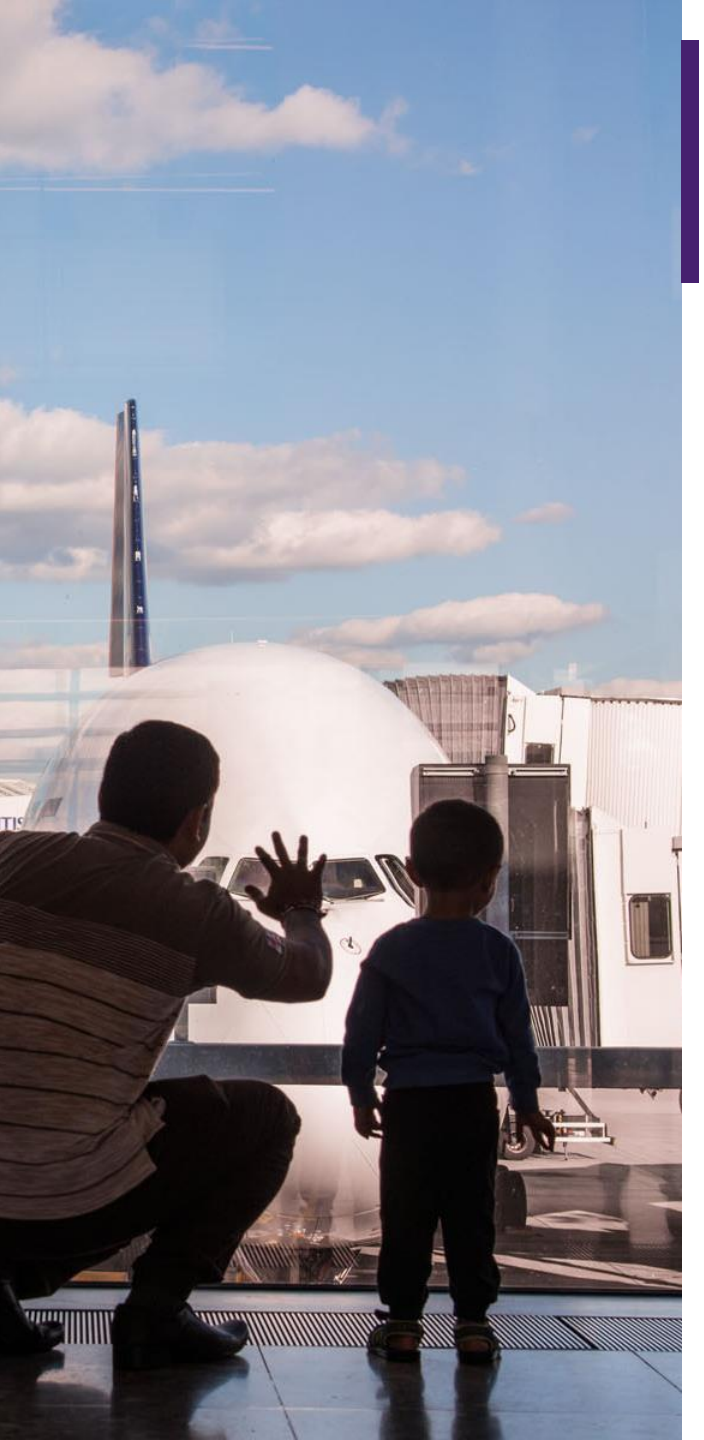
Why Use R?

Challenges Using R

Proposed Solutions

A View Forward

Heathrow

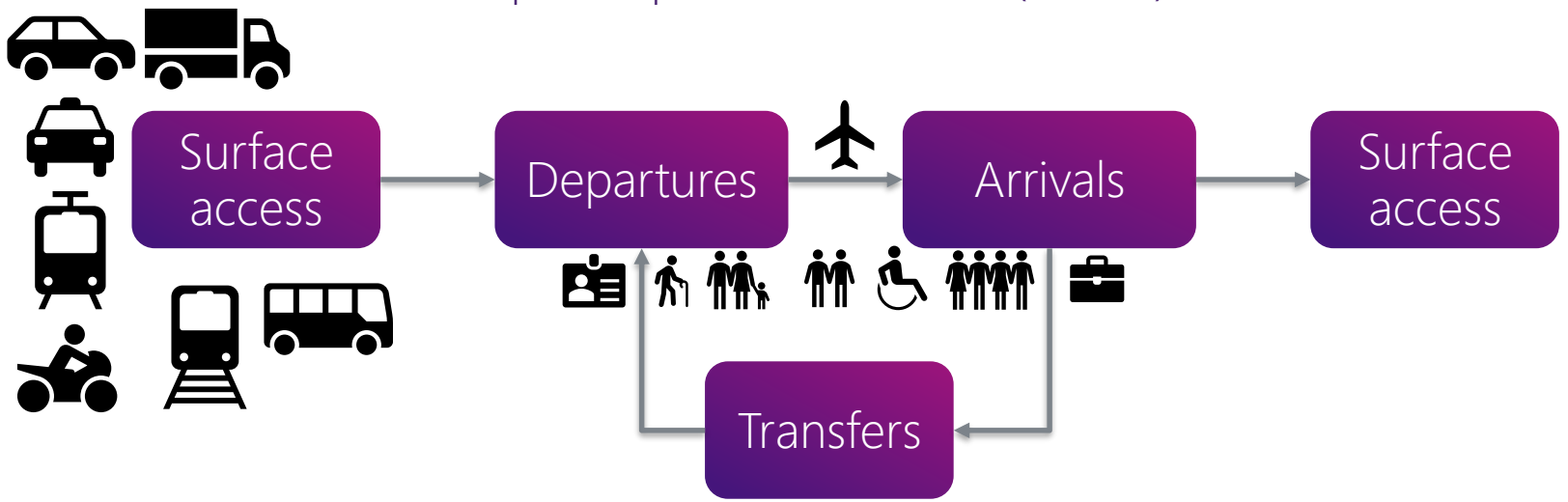


Heathrow Forecasting

	0 to 9 Months	1 to 36 Months	1 to 30+ Years
Forecast	Short Term	Medium Term	Long Term
Purpose	Used for the daily planning and operation of the airport	Used to set budgets and charges, and for initial resource planning	Used for regulation, business planning and long term capacity planning
Inputs	• Latest flight schedules • Passenger data and trends • Airline booking data • Passenger flow data	• Historic and future flight schedules • Passenger data and trends • Short term forecasts	• Economic indicators (GDP, oil price etc) • Passenger data • Medium term forecasts
Outputs	• Arrivals and departures • Total, direct and transfer passengers • Movements • Totals at flight level • 15min flow profiles for security	• Arrivals and departures • Total, direct and transfer passengers • Movements • Monthly totals by airline and destination	• 2-way passengers • Total, direct and transfer passengers • Movements • Annual totals by market

Where is forecast data used?

Airport Operations Centre (APOC)



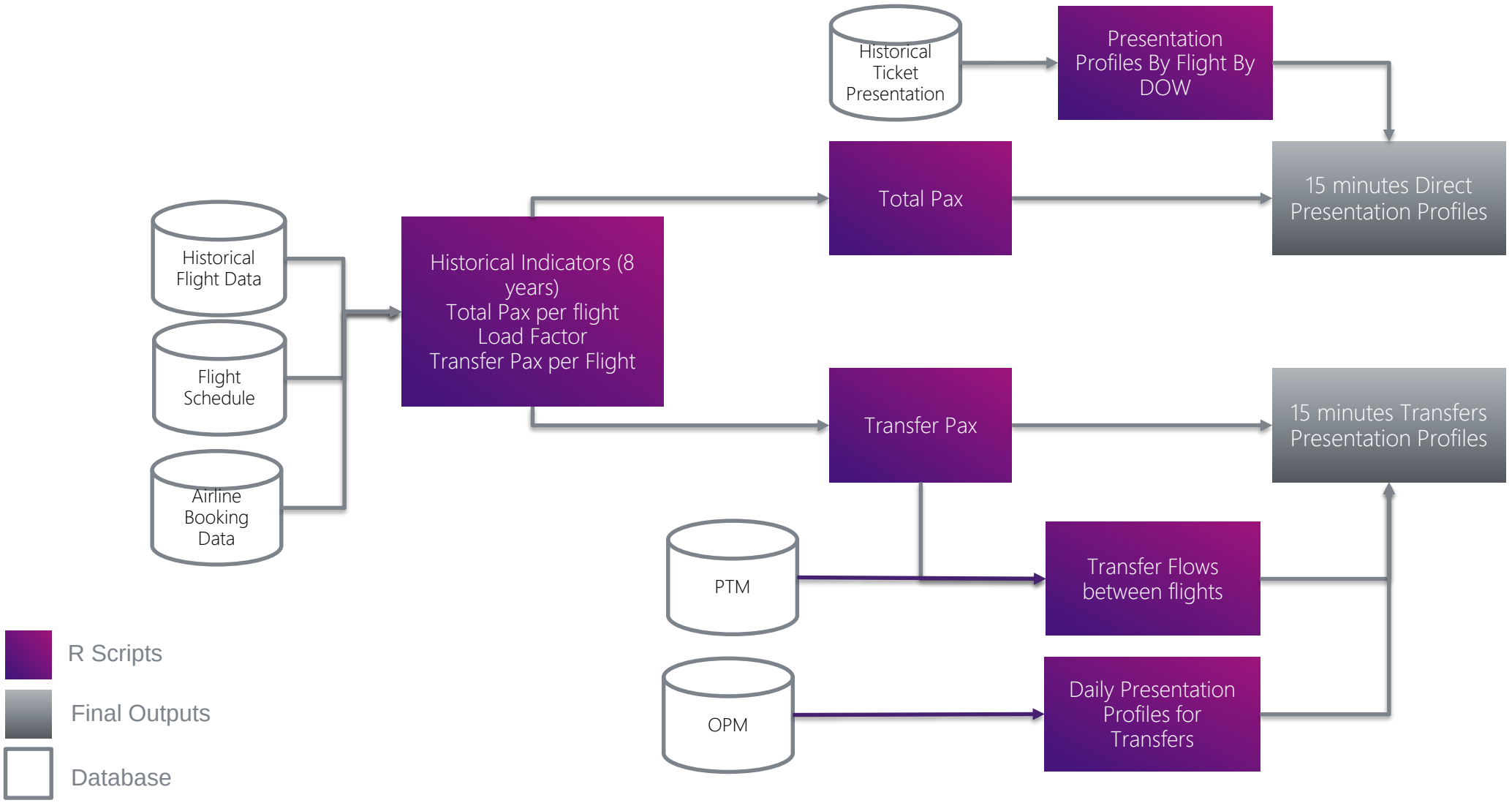
- Surface access team
- Police
- Retail
- Heathrow CRS
- Heathrow Engineering

- Check-in desks
- Security planning
- ITO bussing
- Baggage
- PRM
- Retail
- Engineering
- Cleaning companies
- Capacity planning (up-gage)
- Border Control

Why are accurate forecasts important?

Aim	Why
Improve accuracy	• Real forecasts used across Heathrow • Inaccuracy costs money and passenger satisfaction
Expand scope	• Models developed and built by the team • Constantly evolving to meet needs of customers • Expand outputs to be used in more areas
Increase visibility	• No more "black boxes" – transparent models • Forecasts shared with more people to improve planning across the airport • Improving accessibility through PowerBI reports

Forecasting Process



Tools & Technologies

1. **R**
   Provides state-of-the-art methods for data science
2. **Microsoft AzureML**
   Provides the infrastructure to run scripts in the cloud
3. **MS SQL - SQLite**
   Database systems
4. **PowerBI Dashboards**
   Interactive reports for data sharing

Heathrow

Why use R?

1. **R Is Data Science For Non-Computer Scientists**
   R is a statistical programming language developed by scientists that has open source libraries for statistics, machine learning, and data science. R lends itself well to business because of its depth of topic-specific packages and its communication infrastructure

2. **State-of-the-art functions and lots of documentation**
   R has now one of the richest ecosystems to perform data analysis. There are around 12000 packages available in CRAN. It is possible to find a library for whatever the analysis you want to perform..

3. **Easy to extend, Modify and improve with add-on packages**
   R programming language is open source, highly extensible and easy to learn so developments in R happen at a rapid scale and the community of developers is huge.

Heathrow

n: Confidential

Why use R?

4 **Visualization infrastructure**
R is an amazing platform for data analysis, capable of creating almost any type of graph.

5 **R Shiny**
Shiny is a web application framework for R. In a nutshell this means that anyone who knows some R can start to build applications that sit in a web browser. It could be as simple as displaying some graphics and tables, to a fully interactive dashboard. The important part is that it is all done with R; there are no requirements for web developers to get involved.

6 **Microsoft Support and infrastructure**
R is integrated in many Microsoft apps such as PowerBI and AzureML which are used throughout the forecasting process to generate and share results.

Heathrow

Challenges Using R

1. **Memory Management**
   R commands give little thought to memory management, and so R can consume all available memory.
2. **R does not impose strict rules.**
   One needs a lot of discipline to maintain a proper coding standard. Lack of discipline can quickly lead to a hard-to-maintain R code. Especially, once the code grows bigger.
3. **Little consistency in the algorithms**
   The algorithms developed and their documentation differ from one package to the other making it difficult to adopt an algorithm to your problem.
4. **Time Management**
   Packages written in R language tend to be slower than those written in Python or Matlab.

Proposed Solutions…

Memory Management

- Memory Usage Monitoring — rm(var)
- Garbage Collector — gc()
- Modification In Place — data.table :=

R does not impose strict rules

- Impose style guides!
- Consistent package usage

Time Management

- Vectorization
- Packages instead of base - data.table, strinr
- Parallel code execution — parallel, foreach

Little consistency in the algorithms

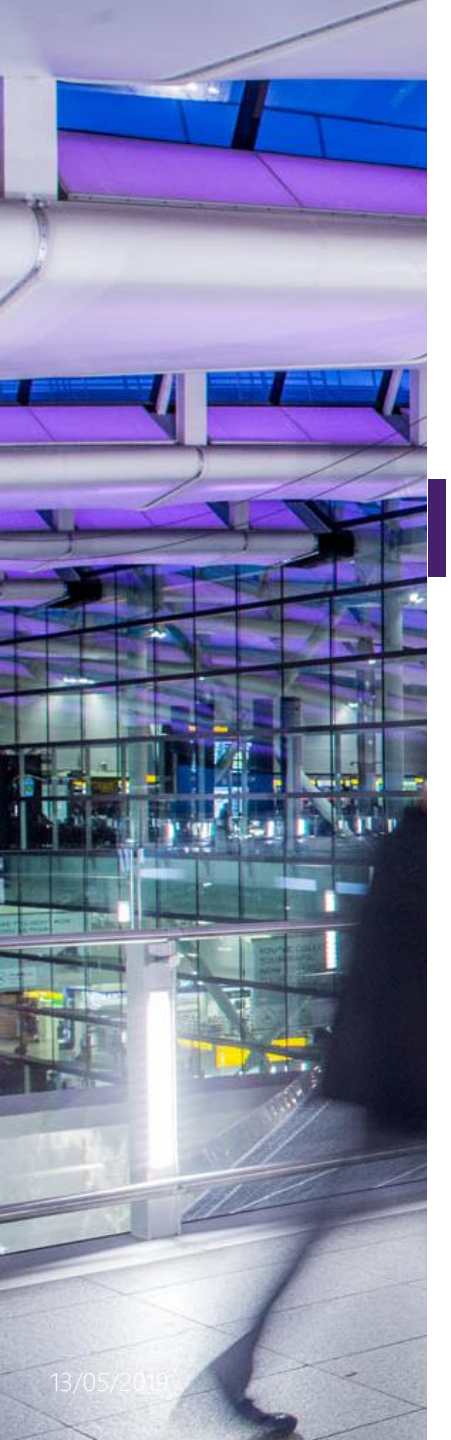
- Consistent package usage

A View Forward…

1. Faster Cloud-Based Implementations
   - More forecast updates
   - Chances for optimization

2. Forecast Scope Expansion
   - Forecast updates for more areas / functions

Heathrow



Thank you for your attention!

Any questions?

Heathrow