Data-driven museums

Alice Daish
Data Scientist
@alice_data

12th International Digital Curation Conference
February 2017
The British Museum

Hardware

Software

Data

@alice_data
Data Silos
Unmanaged data
Undervalued data

The British Museum
The British Museum

Data opportunity

@alice_data
The British Museum
Here to help us ...

Questions

Problems

Data

@alice_data
...to the rescue

Questions

Problems

@alice_data
The British Museum

Questions

Problems

Answers

@alice_data
However…
<table>
<thead>
<tr>
<th>Questions</th>
<th>Data</th>
<th>Answers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Problems</td>
<td></td>
<td>Decision</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Actions</td>
</tr>
<tr>
<td></td>
<td></td>
<td>More Q’s</td>
</tr>
</tbody>
</table>
The British Museum

@alice_data
Data-driven
Data-Driven
Data-Driven = Data in the hands of decision makers
Data-Driven = Data in the hands of decision makers

@alice_data
Data-Driven = Data in the hands of decision makers + De-siloed, accessible, centralised data

@alice_data
Data Science at The British Museum
The British Museum

Set up

Opened 1759 to all 'studious and curious persons'
1st National Public Museum in the World

Today
One of the most visited museum in the world
8 million objects
2 million years of human history
Starting point

Didn’t have …
- No list of data sources
- No data access
- No databases
- No data warehouse

Did have…
- R
- Data Scientist
- Big Data: Senior Product Manager
- What does “big data” mean to the museum?

@alice_data
Finding data
What is a data source? From a Excel spreadsheet to large database e.g. ODIN
Bubble Envy

D3 Bubble Chart: https://bl.ocks.org/mbostock/4063269
The British Museum

British Museum Data Audit 2015
The big data team conducted a survey across the museum and found more than 250 data sources. This visualisation demonstrates the variety of data across the museum.
D3 (Data-Driven Documents)

2017 update
In Development

@alice_data
D3
(Data-Driven Documents)

2017 update
In Development

@alice_data
Business Problems = Data Opportunities

We don’t know who our visitors are?
Online = > 9 million
Offline = 6.8 million

We don’t know what they do in the museum?

And we don’t know the opportunities to generate revenue?
We don’t know who our visitors are?
Online = > 9 million
Offline = 6.8 million

We don’t know what they do in the museum?
And we don’t know the opportunities to generate revenue?

“silos” and “wrangling”
data viz
visitor movement
predictive modelling
“silos” and “wrangling”

Multiple visitor data platforms

CSV exports from external platforms

No SQL
“silos” and “wrangling”

Multiple visitor data platforms
CSV exports from external platforms
No SQL

100’s of columns

print format exports
nested by timeslots

Split first and second name

@alice_data

Packages: data.table, dplyr
No SQL = data.table

How many visitors are on multiple platforms?

Assess the visitor data siloing

Why?
*To improve engagement and access of the museum we need to examine our visitor data.*
Top 500 website search
June 2015-16
Top 500 website search
June 2015-16

The British Museum

Packages: RSiteCatalyst (Adobe Analytics), WordCloud

Visitor Movement

62 galleries, 3 floors, largest covered public square in Europe with 6.8 million visitors per year.
Visitor Movement

Wi-Fi presence used to sample visitor numbers

1st to use R to connect to CISCO Presence API
Visitor Movement

Wi-Fi presence data is used to sample visitor numbers. 1st to use R to connect to CISCO Presence API.

Packages:
- httr
- jsonlite

```
# --------------CISCO API Connection to collect CISCO data-------------
DATE : 03/06/2016
AUTHOR: Alice Balsh, alice@britishmuseum.org

#install and load packages
library(httr)
library(jsonlite)

#load password and username code file
source("loginCisco.R")

# FIND THE LIST OF SITES-------
sites<-GET("https://cxw2cisco.com/api/config/v1/sites", authenticate(user, password))
"get the URL api content including authorization"

# testing different export formats
str(content(sites)) # see content
sitelist<-content(sites, "text") # collects content as text string
sitelist<-fromJSON(sitelist) # convert to table format from string
head(sitelist) # see the top of the table

# List of site name and siteID
sitId<-cbind(sitelist@asU64String, sitelist@name)

# EXAMPLE COLLECT HOURLY DATA OF ALL SITE FOR ONE DAY (14/05/2016)
hourdata<-matrix(NA, nrow = 97, ncol = 27) # blank matrix
colnames(hourdata)<-c("siteId", "siteName", "date", "2016-05-14")
hourly<-GET(paste0("https://mxny1.cisco.com/api/presence/v1/visitor/hourly?siteId=",
siteId[[1]], 
&date=2016-05-14"), authenticate(user, password))

hourdata[1,1]<-siteId[1,1] # ID
hourdata[1,2]<-siteId[2,1] # Name of site
hourdata[1,3]<"2016-05-14"
hourdata[1,4]<content(hourly)[1] 0
hourdata[1,5]<content(hourly)[1] 1
hourdata[1,6]<content(hourly)[2] 2
hourdata[1,7]<content(hourly)[5] 3
```
Visitor Movement

Wi-Fi presence data was used to sample visitor numbers.

1st to use R to connect to CISCO Presence API

@alice_data

Packages: httr, jsonlite
Nothing is perfect
Can we predict ticket sales for exhibitions?

**mixed effect modelling**

- data wrangling
- modelling
- prediction

Packages: lme4, lubridate, dplyr, data.table, ggplot2, ggthemes
Predictive modelling

First initial model created
Predicted first exhibition sales
Development continues …
Recommendations and Insights

Audio guide
1 in 3 visits start at the Rosetta Stone

And 3 of the top visited stops are not on the “Top 10” tour

Digital Signage
Audio guide starting direction
Promoting exhibition tickets at quiet times

Ticketing Trends
Peak and Off-peak times across exhibitions looking at ticketing capacity
Recommendation for filling quiet timeslots and investigation into time dependent ticketing pricing
De-siloed accessible data
Data Pipeline

- Data Sources
- Data Integrations
- Data transformation
- Data warehouse
- Databases
- Data science tools
- Analytics tools
- Raw data
- Data Lake

NEW insights
Monitoring
Analysis

Dashboards
Reporting
Visualisation
Support staff empowerment

- Data in the hands of decision makers
- De-siloed accessible centralised data

Data-driven decision making primarily for product and content development and optimization

Scoping opportunities for revenue generation

BUT Who knows?

**Internet of things**
e.g. Toilet door locks, Boilers, Visitor Flow Signs

**Machine Learning**
e.g. Predicting Visitor Numbers, Optimization, New Collection Discoveries
Transformation advice
Problems to becoming data-driven
Problems to becoming data-driven
Problems to becoming data-driven

@alice_data
Problems to becoming data-driven

@alice_data
Let data drive decisions, not the Highest Paid Person's Opinion.

#HowGoogleWorks

HowGoogleWorks.net
DATA CULTURE
Transforming any organisation to be data-driven is about changing behaviours and habits.

DATA ADVOCACY
Data value needs to be discovered and shared. By giving voices to data champions and advocates.
support

High level support
Quick Wins
Find early adopters
Build trust
Technical Evangelist
Technology orientated and use data frequently within their roles.
Supports data roles and adoption across the business

Data Ambassador
Data is highly important to their role but not necessarily their main responsibility.
Adopting data techniques and sharing the value and insight data can provide.

Business Data Advocates
Supports data application across business
Champion the use of data in high-level meetings and discussions
team skills

Data Science

- Maths & Statistics Knowledge
- Computer Science

Domain Expertise

Communication & Empathy

@alice_data
team skills

The British Museum

Strategic Project Management
- Stakeholder Management
- Project Management
- Communication & Empathy
- Business Strategy

@alice_data
upskilling
Data Science Upskilling

Beginner
- Excel main tool
  - Filter
  - Basic calculations

Intermediate

Advanced

@alice_data
Data Science Upskilling

Beginner
- Excel main tool
  - Filter
  - Basic calculations

Intermediate
- Excel and BI tool
  - Pivots
  - Advanced Calculation
  - Visualisation
  - Data Analysis Reporting

Advanced
- R

@alice_data
Data Science Upskilling

**Beginner**
- Excel main tool
  - Filter
  - Basic calculations

**Intermediate**
- Excel and BI tool
  - Pivots
  - Advanced Calculation
  - Visualisation
  - Data Analysis Reporting

**Advanced**
- Tools including Programming
  - Data Wrangling/Munging
  - Visualisation
  - Data Analysis
  - Modelling/Machine Learning
Data Science Upskilling

Beginner
- Excel main tool
  - Filter
  - Basic calculations

Intermediate
- Excel and BI tool
  - Pivots
  - Advanced Calculation
  - Visualisation
  - Data Analysis Reporting

Advanced
- Tools including Programming
  - Data Wrangling/Munging
  - Visualisation
  - Data Analysis
  - Modelling/Machine Learning

Excel
Data Science Upskilling

Beginner
- Excel main tool
  - Filter
  - Basic calculations

Intermediate
- Excel and BI tool
  - Pivots
  - Advanced Calculation
  - Visualisation
  - Data Analysis Reporting

Advanced
- Tools including Programming
  - Data Wrangling/Munging
  - Visualisation
  - Data Analysis
  - Modelling/Machine Learning

@alice_data
## Data Science Upskilling

**Beginner**
- Excel main tool
  - Filter
  - Basic calculations

**Intermediate**
- Excel and BI tool
  - Pivots
  - Advanced Calculation
  - Visualisation
  - Data Analysis Reporting

**Advanced**
- Tools including Programming
  - Data Wrangling/Munging
  - Visualisation
  - Data Analysis
  - Modelling/Machine Learning

Excel

BI tools/dashboard

Programming: R/Python
listen & empathy

Listen to your stakeholders
Help them formulate
De-complicate keep it simple

@alice_data
communicate

Comms series “Did you know”

Outreach

Encourage data ambassador

@alice_data
The British Museum

process

The British Museum
Big Data Team : 12 step Analytical Process

1. Identify a problem or question
2. Agree the business questions and/or hypothesis with stakeholders
3. Identify the data available
4. Decide how best to capture the data
5. Decide how to clean, transform and deal with gaps
6. Decide on how the data will be managed, stored, accessed, manipulated and scheduled
7. Analyze
8. Create report or visualization
9. Provide insights & storytelling back
10. Discuss how to embed these results into the business and/or future decision-making processes
11. Discuss how to better capture data in the future (new data or improve existing methods)
12. Can we close the project, or is there a new question?

@alice_data
Vision for the future
55,000 museums 180 countries


Business Problems

= Data opportunity
Many thanks to museum departments for their support and data access, Siorna Ashby, the big data team and my best friend R.