The British Museum

Data-driven museums

Alice Daish
Data Scientist
@alice_data

12th International Digital Curation Conference
February 2017

CC BY NC SA

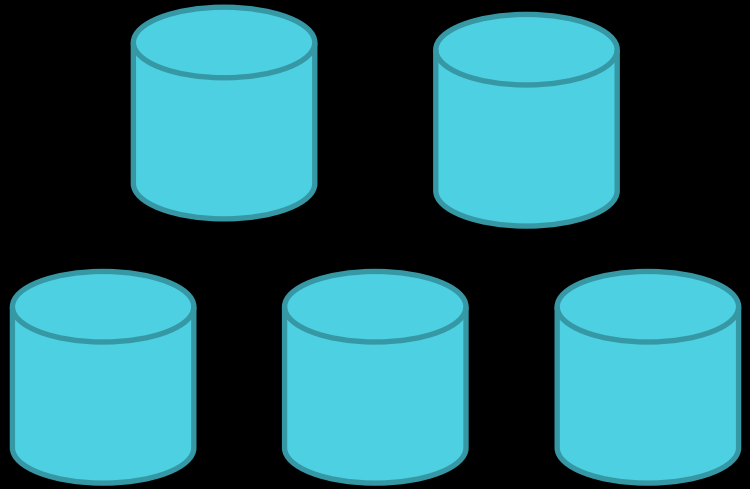
The British Museum

@alice_data

cc BY NC SA

The British Museum

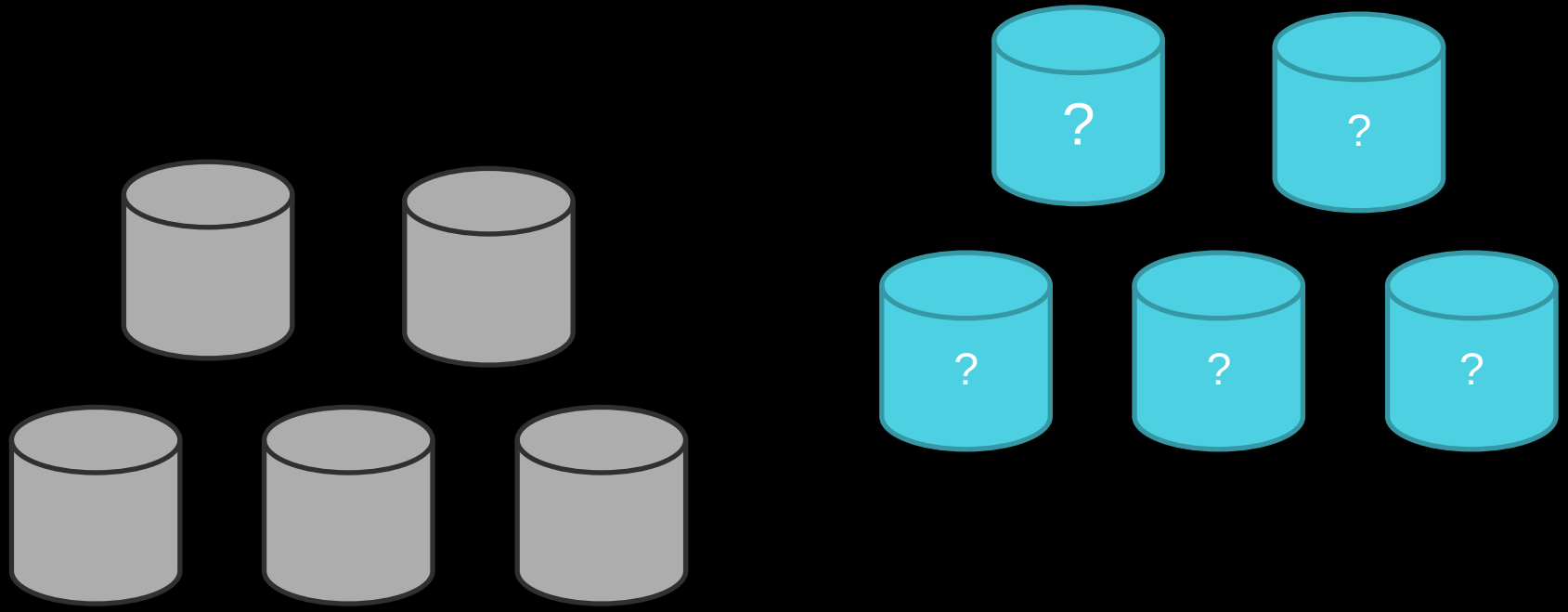
Data Silos



cc BY NC SA

The British Museum

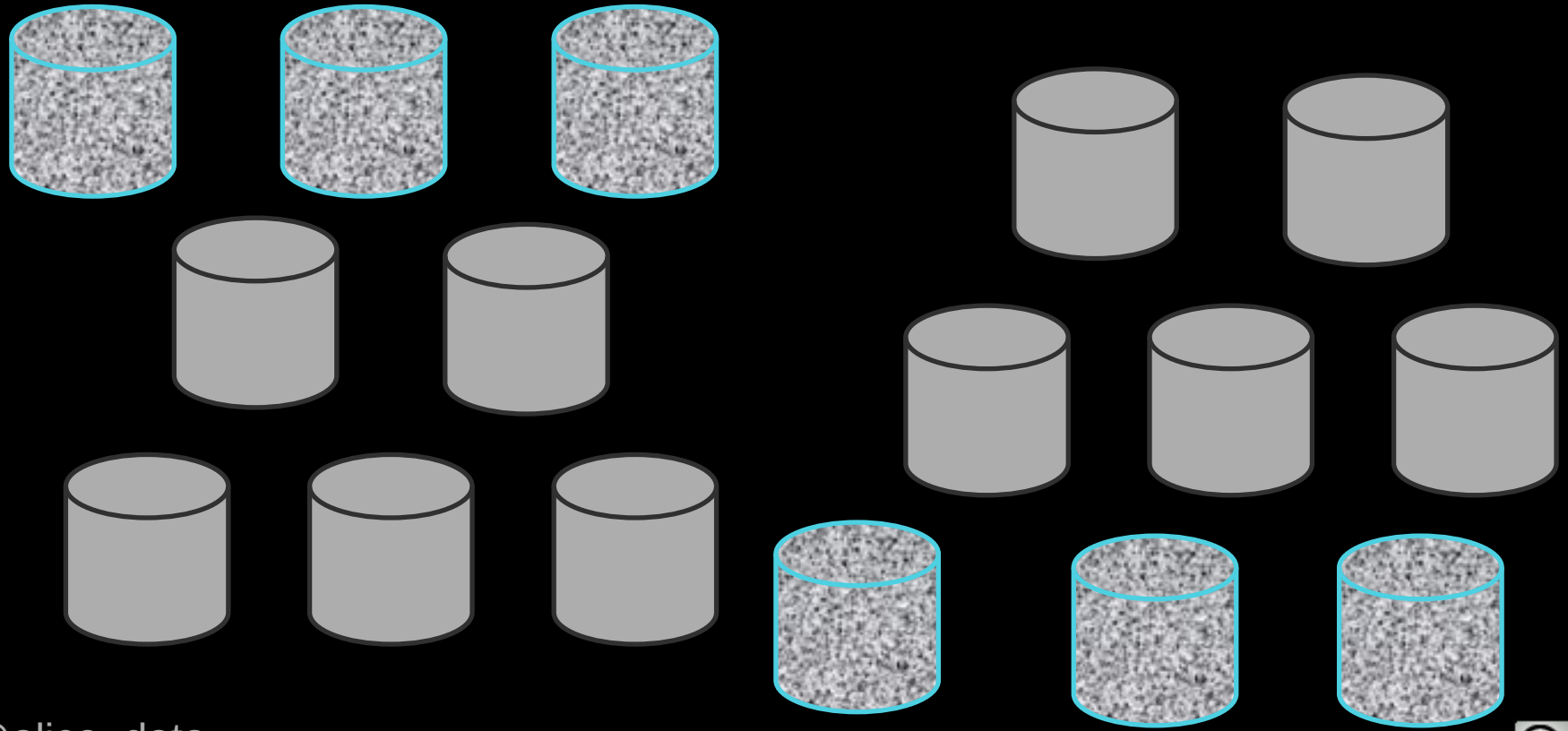
Unmanaged data



cc BY NC SA

The British Museum

Undervalued data

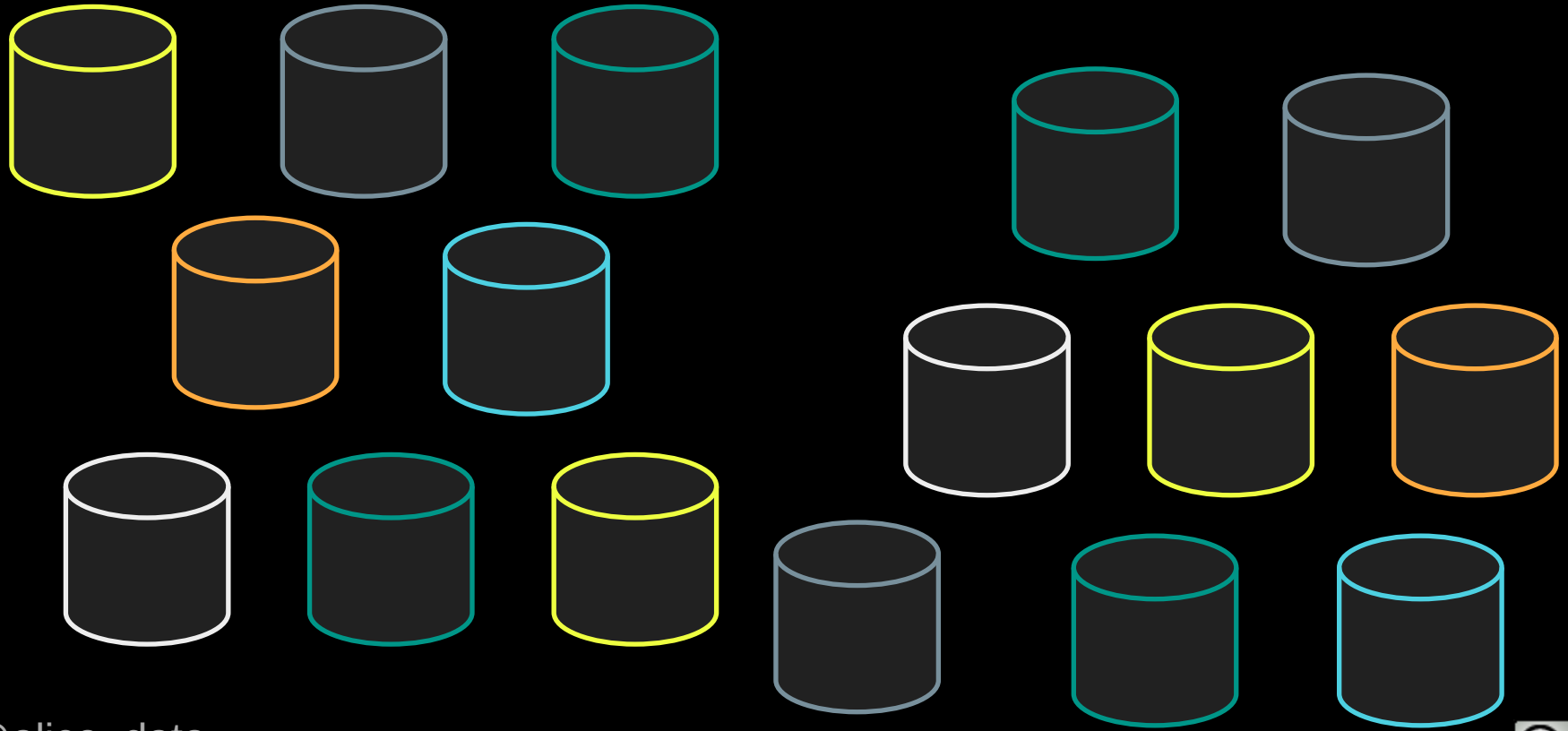


@alice_data

cc BY NC SA

The British Museum

Data opportunity



@alice_data

cc BY NC SA

The British Museum

Questions

Problems

cc BY NC SA

The British Museum

@alice_data

CC BY NC SA

The British Museum

@alice_data

CC BY NC SA

The British Museum

Here to help us …

cc BY NC SA

The British Museum …to the rescue

Questions

Problems

@alice_data

cc BY NC SA

The British Museum

Questions

Problems

Answers

@alice_data

cc BY NC SA

The British Museum

Questions

Problems

Answers

Decision

cc BY NC SA

The British Museum

Questions

Problems

Answers

Decision

Actions

cc BY NC SA

The British Museum

Questions

Problems

Answers

Decision

Actions

More Q's

cc BY NC SA

The British Museum

However…

CC BY NC SA

The British Museum

cc BY NC SA

The British Museum

Questions

Problems

Answers

Decision

Actions

More Q's

cc BY NC SA

The British Museum

@alice_data

CC BY NC SA

The British Museum

@alice_data

cc BY NC SA

The British Museum

Questions	Data	Answers
		Decision
Problems		Actions
		More Q's

@alice_data

CC BY NC SA

The British Museum

Data-driven

BY NC SA

The British Museum

Data-Driven

@alice_data

CC BY NC SA

The British Museum

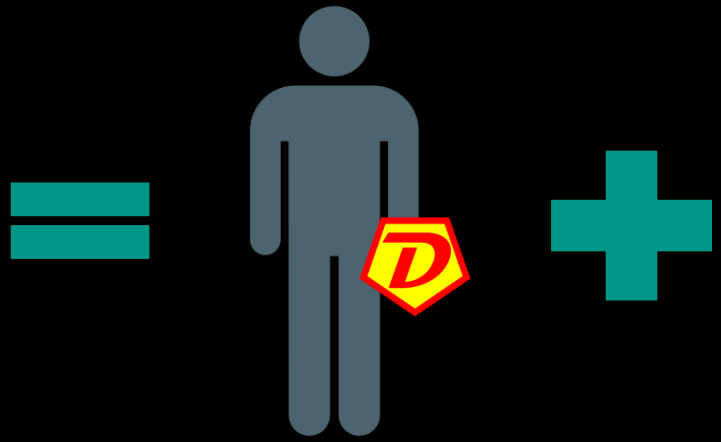
Data-Driven

Data in the
hands of
decision makers

cc BY NC SA

The British Museum

Data-Driven

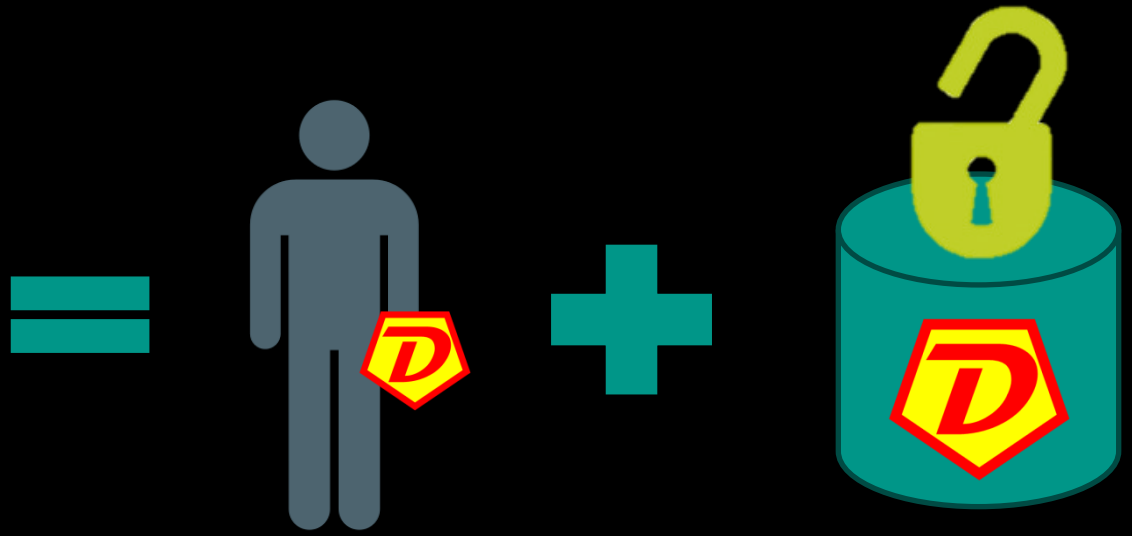


Data in the
hands of
decision makers

cc BY NC SA

The British Museum

Data-Driven



Data in the hands of decision makers

De-siloed, accessible, centralised data

CC BY NC SA

The British Museum

Data Science at The British Museum

cc BY NC SA

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

CC BY NC SA

The British Museum

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?

cc BY NC SA

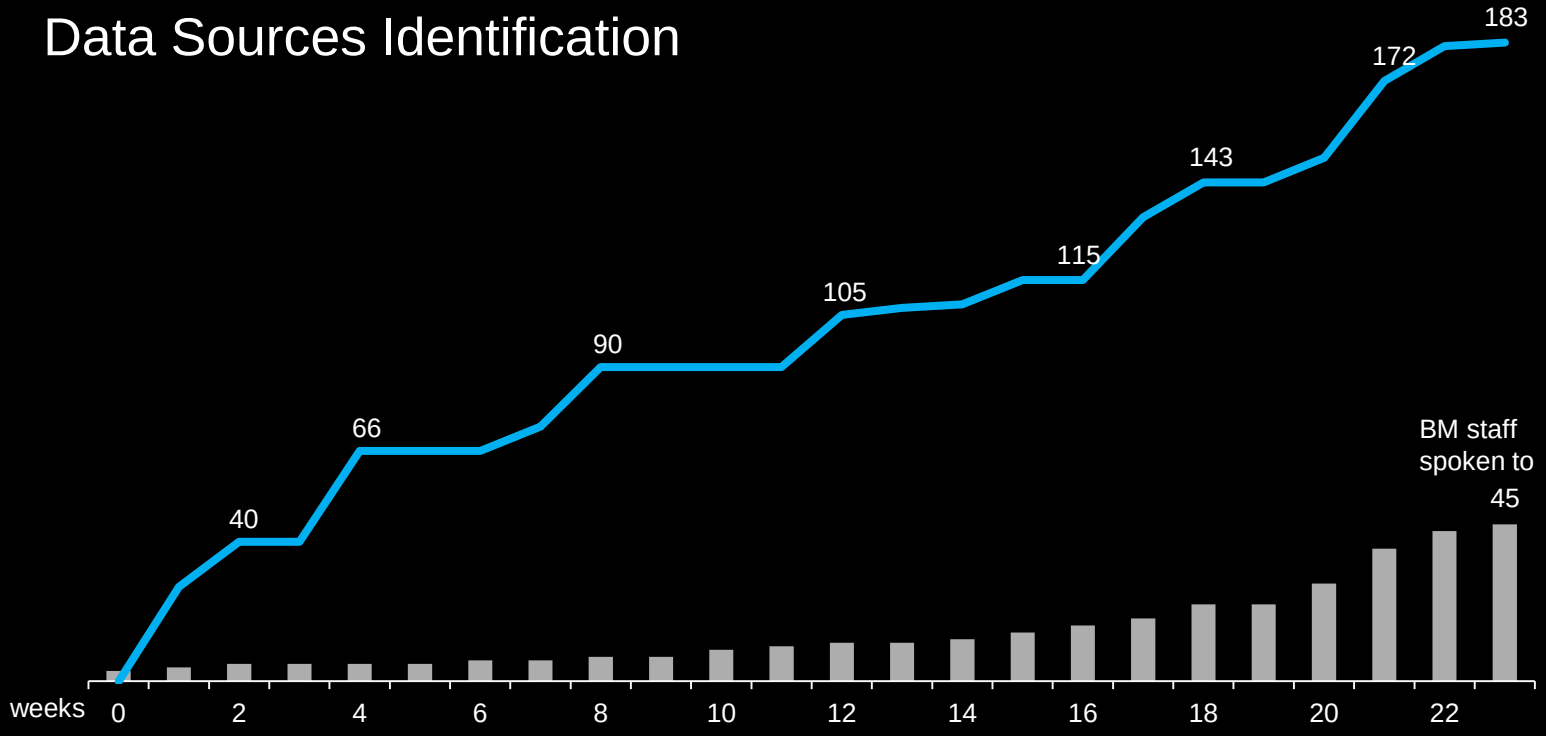
The British Museum

Finding data

CC BY NC SA

The British Museum

Data Sources Identification



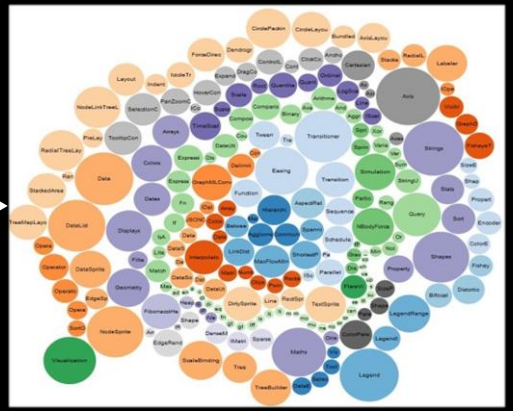
@alice_data

What is a data source? From a excel spreadsheet to large database e.g. ODIN

CC BY NC SA

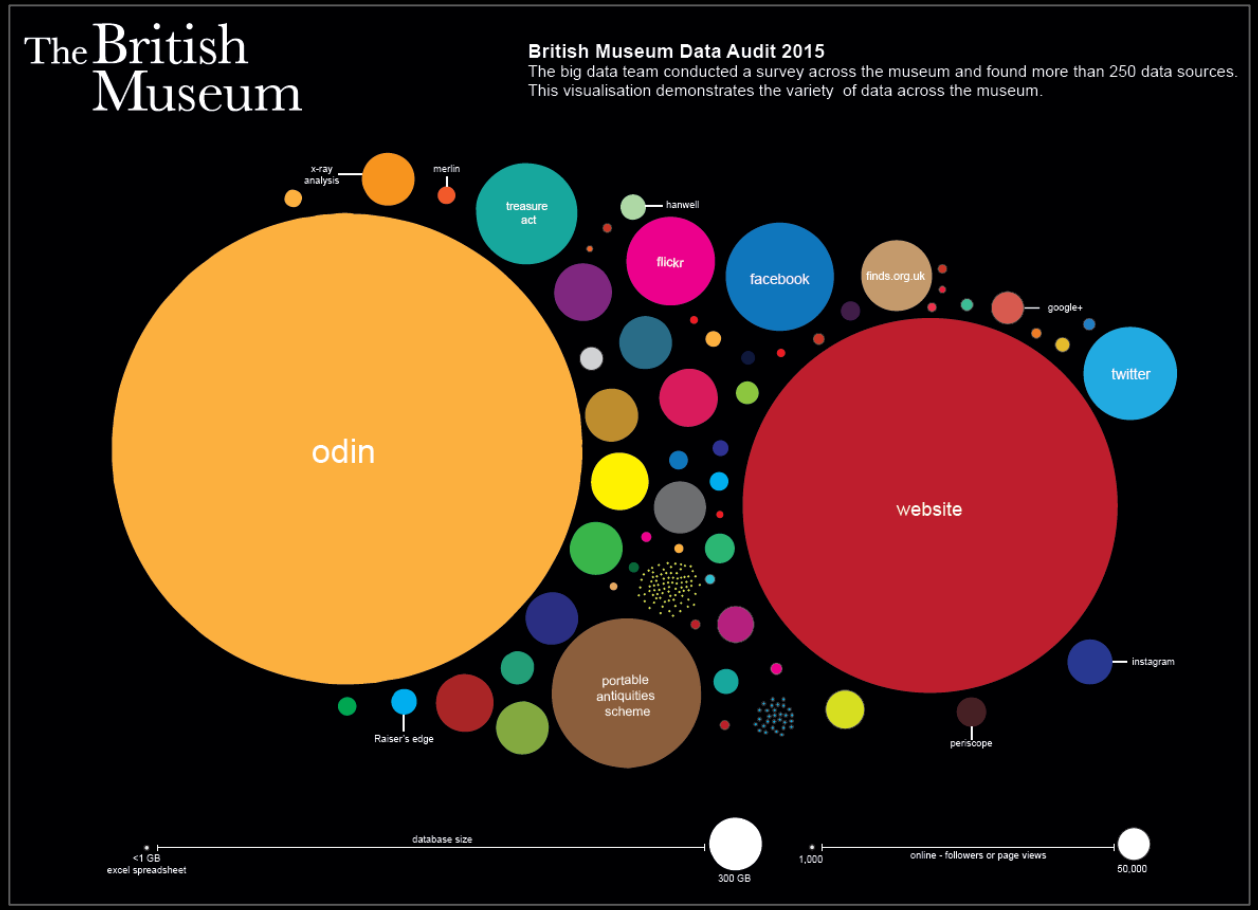
The British Museum

Bubble Envy



cc BY NC SA

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

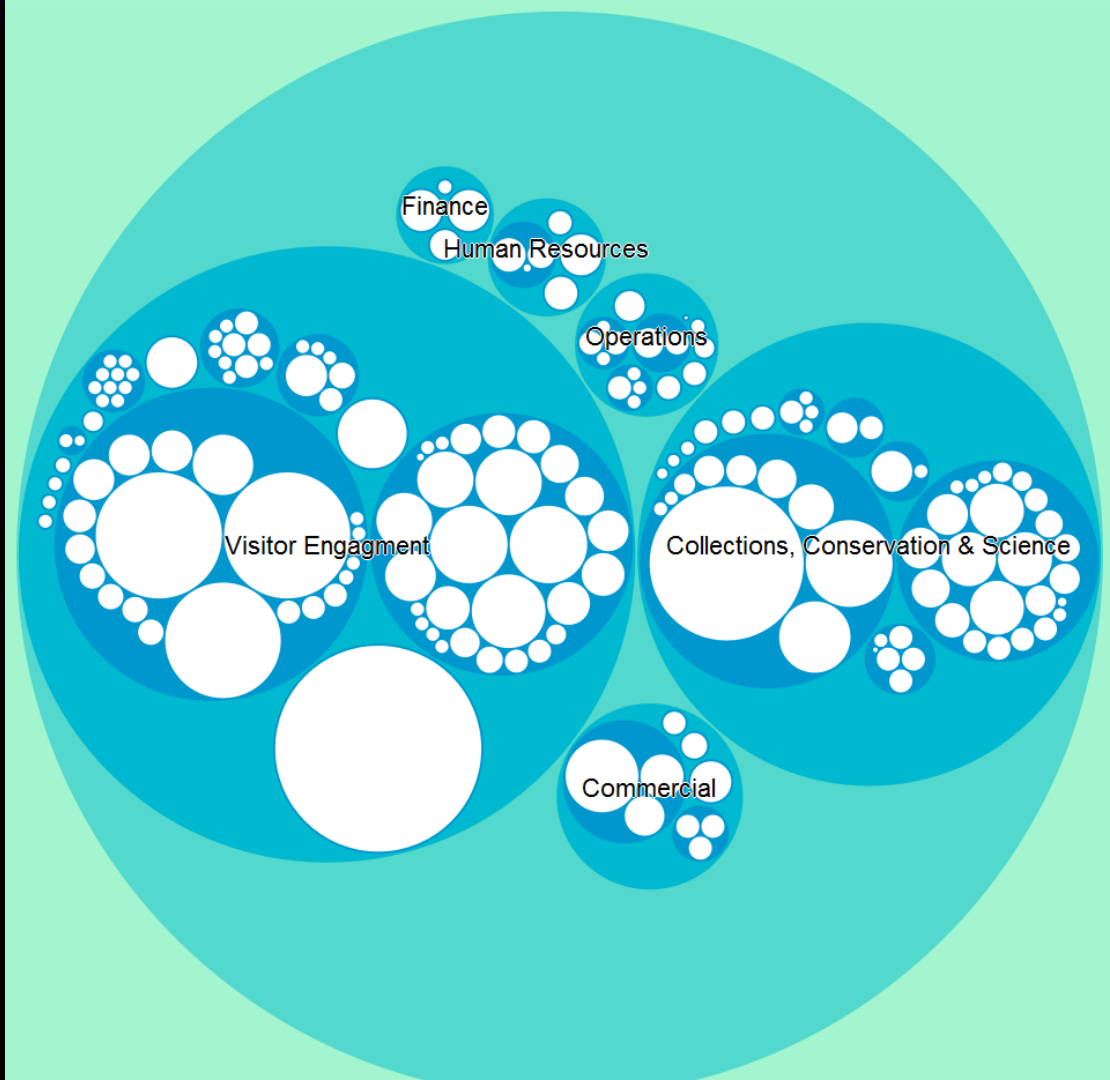


@alice_data

CC BY NC SA

D3

(Data-Driven Documents)



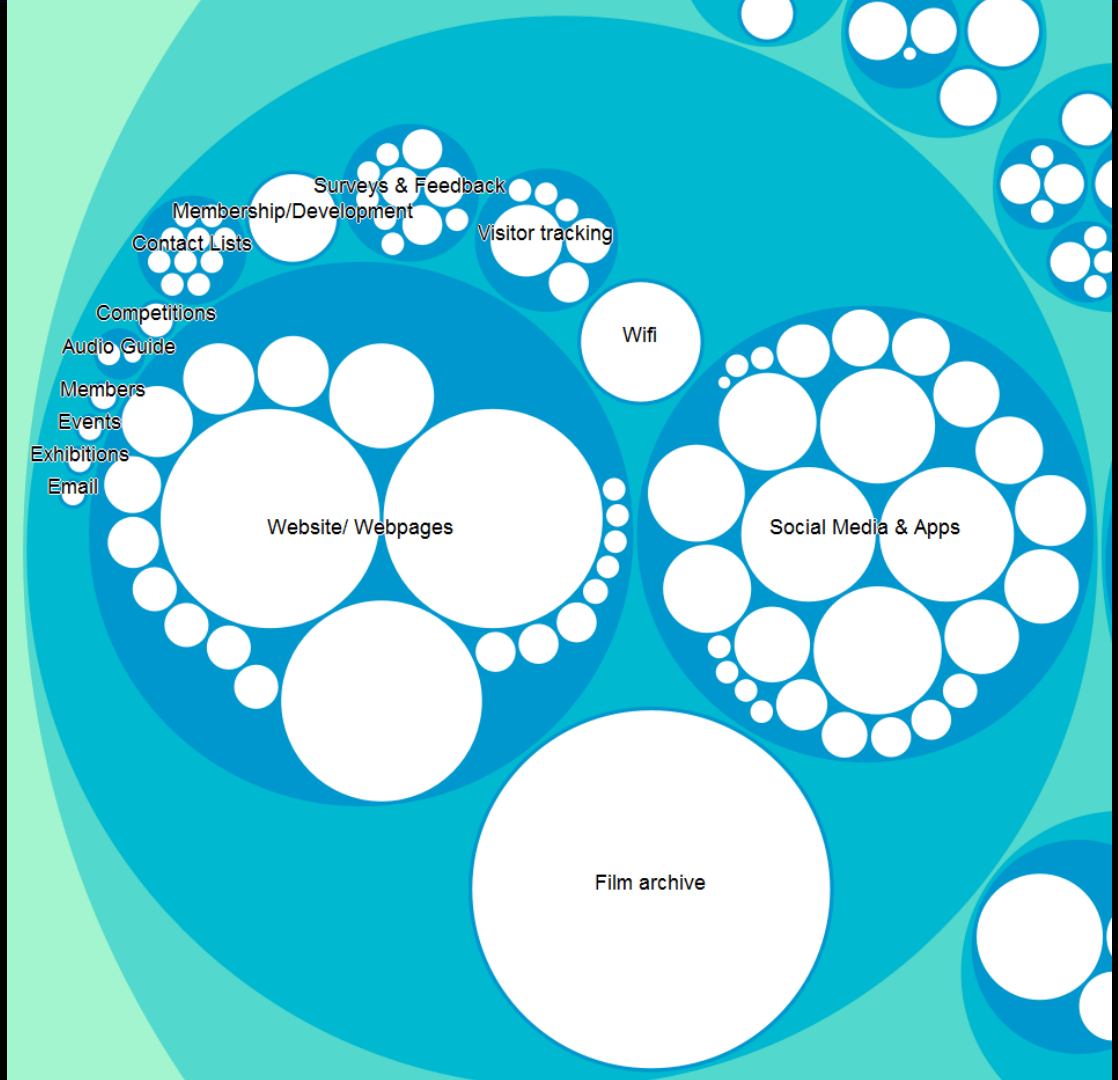
2017 update
In Development

@alice_data

cc BY NC SA

D3

(Data-Driven Documents)



2017 update
In Development

@alice_data

cc BY NC SA

The British Museum

Valuing data

cc BY NC SA

The British Museum

@alice_data

Symbol - Noun Project: Dara Ullrich, DE

The British Museum

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?

cc BY NC SA

The British Museum

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?

"silos" and "wrangling"

data viz

visitor movement

predictive modelling

cc BY NC SA

The British Museum

“silos” and “wrangling”

Multiple visitor data platforms

CSV exports from external platforms

No SQL

cc BY NC SA

Packages: data.table, dplyr

The British Museum

“silos” and “wrangling”

Multiple visitor data platforms

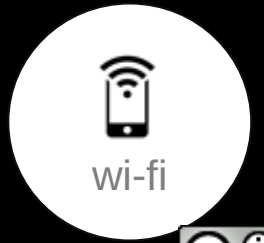
CSV exports from external platforms

No SQL

100’s of columns

print format exports nested by timeslots

Split first and second name



BY NC SA

Packages: data.table, dplyr

The British Museum

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.

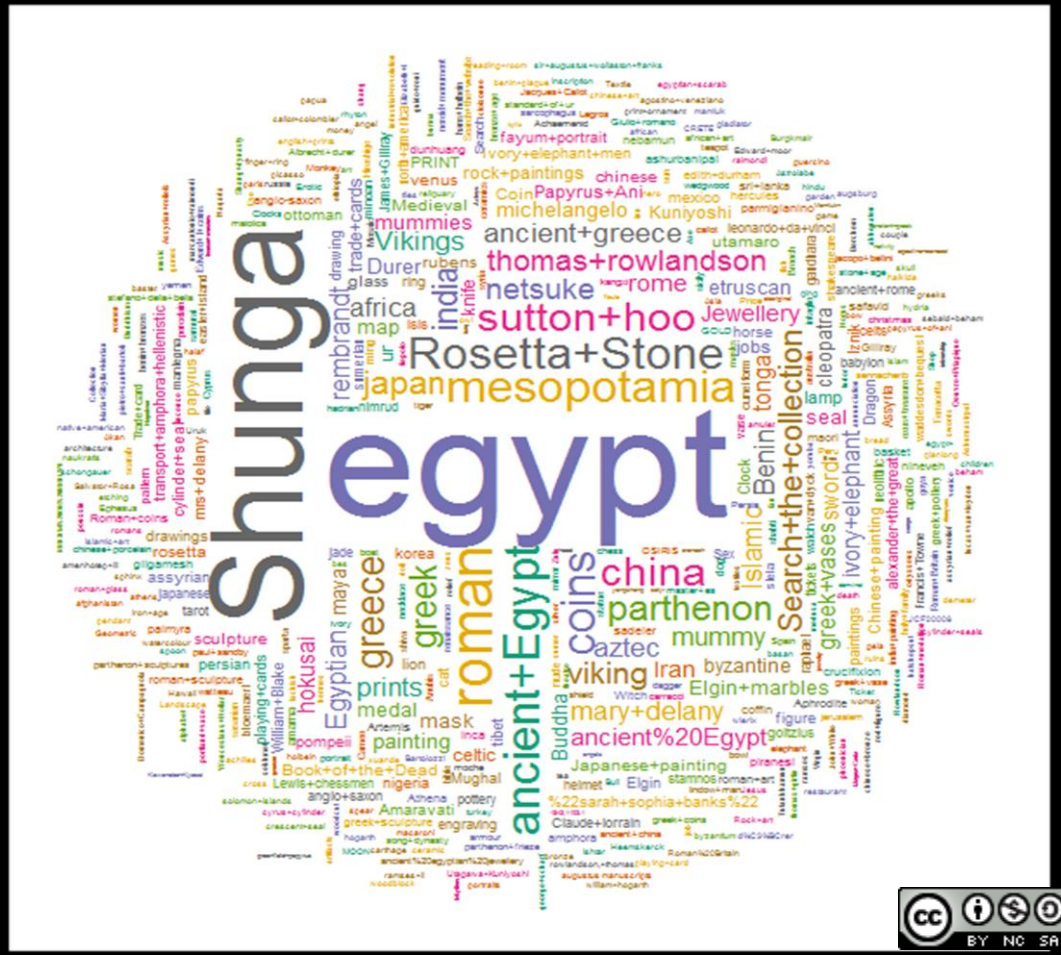
CC BY NC SA

Packages: data.table, dplyr

The British Museum

Top 500 website search

June 2015-16



CC BY NC SA

Packages : RSiteCatalyst (Adobe Analytics), WordCloud

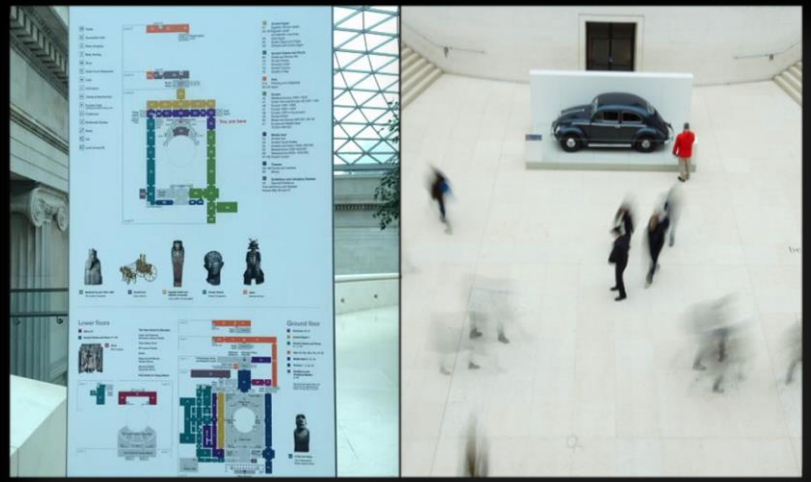
@alice_data

Alice Daish. (2016). R KeyWord Search WordCloud: First Release. Zenodo. 10.5281/zenodo.56283

The British Museum

Visitor Movement

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



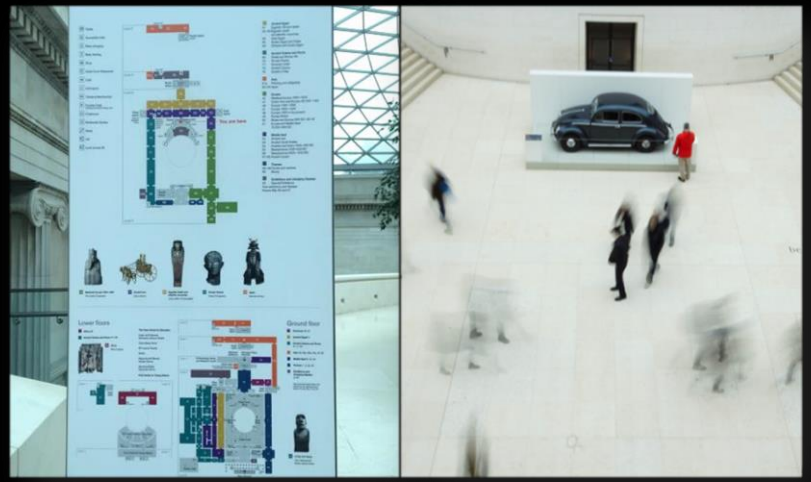
cc BY NC SA

Packages : httr, jsonlite

The British Museum

Visitor Movement

Wi-Fi presence used to sample visitor numbers



1st to use R to connect to CISCO Presence API

CC BY NC SA

Packages : httr, jsonlite

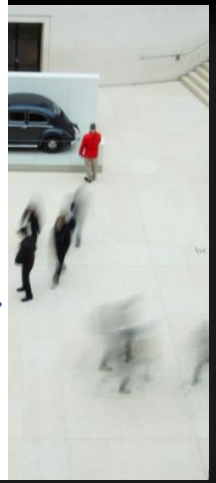
The British Museum

Visitor M

Wi-Fi presenc
used to sampl
visitor numbe

1st to use R to connect to CIS Presence API

```r
#-------------CISCO API Connection to collect CISCO data--------------
#DATE : 03/06/2016
#AUTHOR: Alice Daish adaish@britishmusuem.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://cmxcisco.com/api/config/v1/sites",authenticate(user, password))
# gets the URL api content including authroziation

#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
siteId<-cbind(sitelist$aesUidString,sitelist$name)

#EXAMPLE COLLECT HOURLY DATA OF ALL SITE FOR ONE DAY (14/05/2016)
hourdata<-matrix(NA,nrow = 97*1, ncol = 27) #blank matrix
colnames(hourdata)<-c("SiteID","SiteName","Date","0","1","2","3","4","5",
                      "6","7","8","9","10","11","12","13","14","15","16",
                      "17","18","19","20","21","22","23") #Label columns

  hourly<-GET(paste0("https://1nzgy2.cmxcisco.com/api/presence/v1/visitor/hourly?siteId=",
                     siteId[i,1],"&date=2016-05-14"),authenticate(user, password))

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



CC BY NC SA

Packages : httr, jsonlite

The British Museum

Visitor M

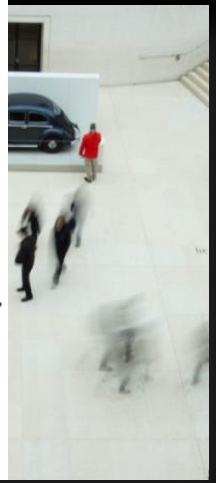
Wi-Fi presenc
used to sampl
visitor numbe

1st to use R to
connect to CIS
Presence API

```
#-------------CISCO API Connection to collect CISCO data--------------
#DATE : 03/06/2016
#AUTHOR: Alice Daish adaish@britishmusuem.org
```

9-10am	10-11am	11am-12p	12-13pm	13-14pm	14-15pm	15pm-16p	16-17pm	17-18pm	Total
23	155	212	254	174	165	155	141	120	1399
38	76	94	153	120	115	101	158	92	947
26	52	82	92	91	84	80	105	54	666
27	50	68	81	74	65	83	87	98	633
7	23	64	93	68	65	69	85	36	510
11	61	78	61	76	48	61	57	37	490
13	56	45	88	47	27	31	78	69	454
26	40	42	65	56	58	48	47	48	430
14	29	53	111	48	31	35	72	35	428
1	26	69	62	51	64	66	56	27	422
1	39	78	57	52	45	53	52	19	396
0	30	56	57	53	59	59	53	23	390
3	24	56	56	47	55	41	43	15	340
15	18	29	30	34	41	51	36	86	340
14	43	30	34	44	43	36	31	50	325
5	20	48	73	38	36	30	44	27	321
7	39	53	43	38	30	42	39	27	318
3	22	44	54	41	36	39	50	26	315
4	16	39	50	50	38	38	42	36	313
0	30	31	28	43	44	37	38	16	267
6	17	30	53	24	24	31	40	28	253
19	27	20	30	18	22	23	31	38	228
2	14	21	39	37	35	25	20	22	215
8	27	32	23	27	29	25	18	22	211
12	12	15	22	28	27	29	25	37	207
22	38	33	24	22	19	15	16	15	204
11	12	19	33	32	21	25	32	17	202
2	17	26	26	31	26	25	30	12	195
10	33	22	25	10	17	20	29	4	170
2	19	17	29	31	19	20	19	12	168
0	9	23	28	19	17	25	35	11	167
0	15	24	31	13	16	20	35	12	166
11	14	21	33	23	20	18	13	8	161
11	9	27	28	23	23	19	12	6	158
3	19	19	25	15	15	24	18	18	156

```
hourdata[i,7]<-content(hourly)$`3`
```



CC BY NC SA

Packages : httr, jsonlite

The British Museum

Nothing is perfect

CC BY NC SA

The British Museum

Predictive modelling

Can we predict ticket sales for exhibitions?

mixed effect modelling

- data wrangling
- modelling
- prediction

CC BY NC SA

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

The British Museum

Predictive modelling

First initial model created
Predicted first exhibition sales
Development continues …

CC BY NC SA

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

The British Museum

Recommendations and Insights

Audio guide

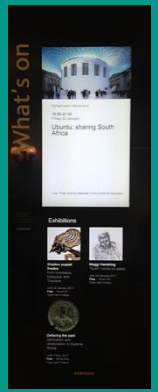
1 in 3 visits start at the Rosetta Stone

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

CC BY NC SA

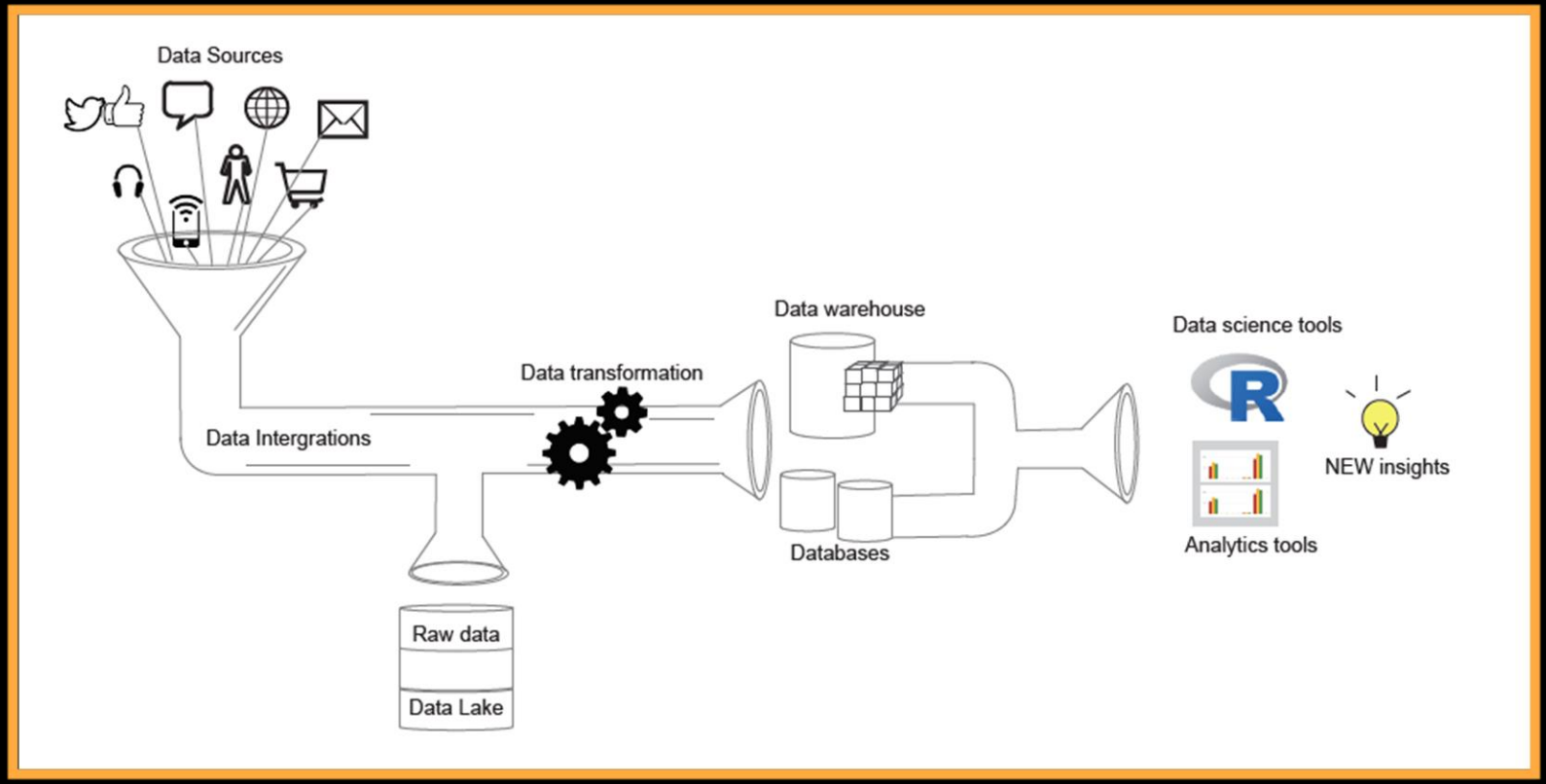
The British Museum

De-siloed accessible data

cc BY NC SA

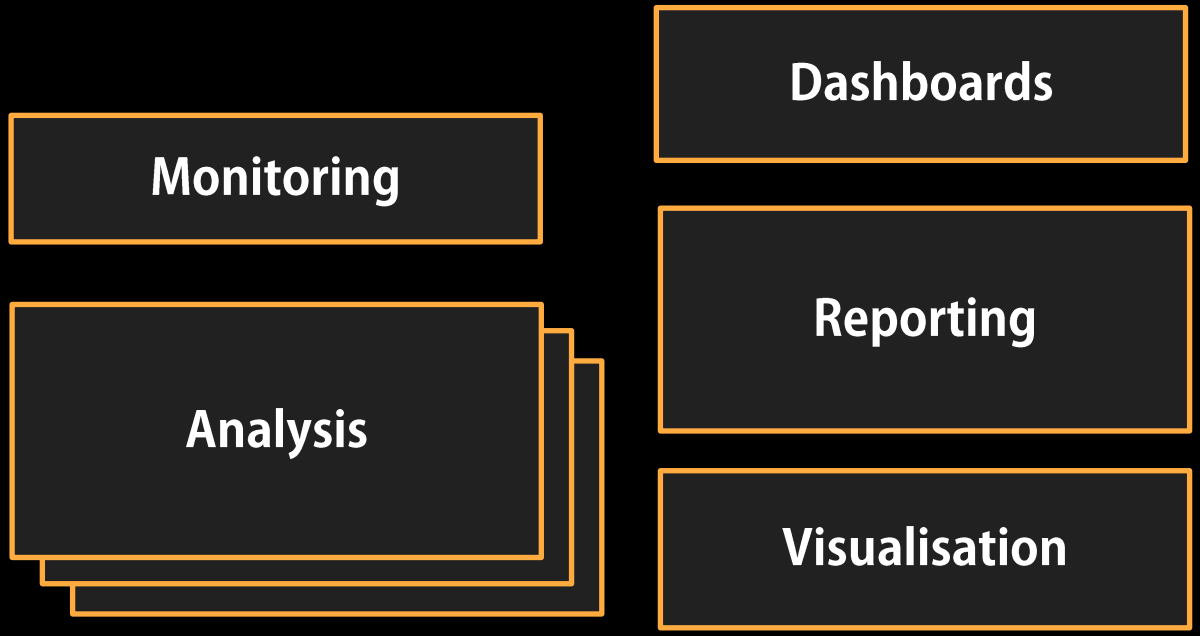
The British Museum

Data Pipeline



CC BY NC SA

The British Museum



cc BY NC SA

The British Museum

What next…

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

Data-Driven

cc BY NC SA

The British Museum

Transformation advice

cc BY NC SA

The British Museum

Problems to becoming data-driven

cc BY NC SA

The British Museum

Siloed and inaccessible data

=

Problems to becoming data-driven

CC BY NC SA

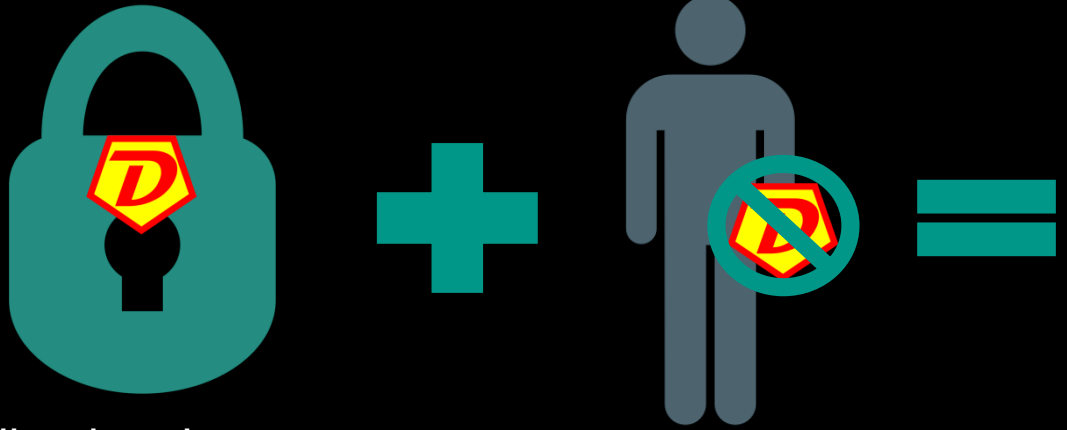
The British Museum

Siloed and inaccessible data

Problems to becoming data-driven

CC BY NC SA

The British Museum



Siloed and inaccessible data

Data not in the hands of decision makers

Problems to becoming data-driven

@alice_data

CC BY NC SA

The British Museum

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.

cc BY NC SA

The British Museum

support

High level support

Quick Wins

Find early adopters

Build trust

cc BY NC SA

The British Museum

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

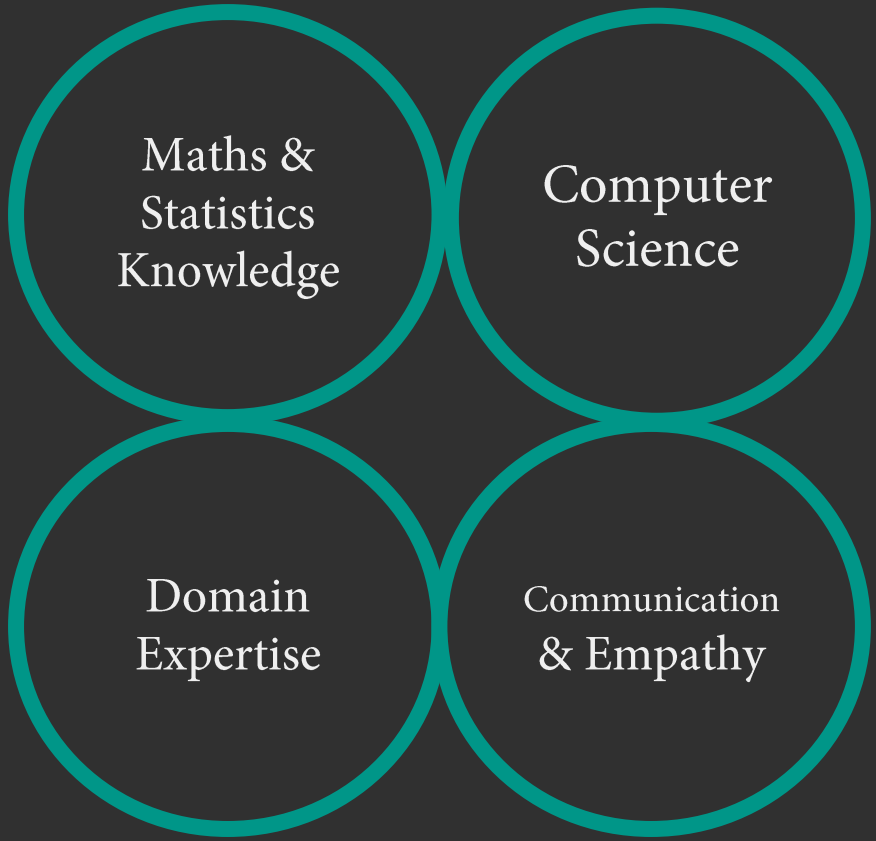
Champion the use of data in high-level meetings and discussions

CC BY NC SA

The British Museum

team skills

Data Science



@alice_data

CC BY NC SA

The British Museum

team skills

Strategic Project Management

- Stakeholder Manage-ment
- Project Manage-ment
- Communication & Empathy
- Business Strategy

@alice_data

cc BY NC SA

The British Museum

upskilling

cc BY NC SA

The British Museum

Data Science Upskilling

cc BY NC SA

The British Museum

Data Science Upskilling

Excel main tool

- Filter
- Basic calculations

CC BY NC SA

The British Museum

Data Science Upskilling

Beginner	Intermediate	Advanced

Excel main tool
- Filter
- Basic calculations

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

@alice_data

CC BY NC SA

The British Museum

Data Science Upskilling

Beginner	Intermediate	Advanced

Excel main tool
- Filter
- Basic calculations

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

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

@alice_data

CC BY NC SA

The British Museum

Data Science Upskilling

Beginner	Intermediate	Advanced
Excel main tool • Filter • Basic calculations	Excel and BI tool • Pivots • Advanced Calculation • Visualisation • Data Analysis Reporting	Tools including Programming • Data Wrangling/Munging • Visualisation • Data Analysis • Modelling/Machine Learning

Excel

The British Museum

Data Science Upskilling

Beginner	Intermediate	Advanced
Excel main tool • Filter • Basic calculations	Excel and BI tool • Pivots • Advanced Calculation • Visualisation • Data Analysis Reporting	Tools including Programming • Data Wrangling/Munging • Visualisation • Data Analysis • Modelling/Machine Learning

Excel

BI tools/dashboard

The British Museum

Data Science Upskilling

Beginner	Intermediate	Advanced

Excel main tool
- Filter
- Basic calculations

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

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

Excel

BI tools/dashboard

Programming : R/Python

@alice_data

The British Museum

listen & empathy

Listen to your stakeholders

Help them formulate

De-complicate keep it simple

cc BY NC SA

The British Museum

communicate

Comms series “Did you know”

Outreach

Encourage data ambassador

@alice_data

CC BY NC SA

The British Museum

process

@alice_data

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 existng methods)
12		Can we close the project, or is there a new question?

CC BY NC SA

The British Museum

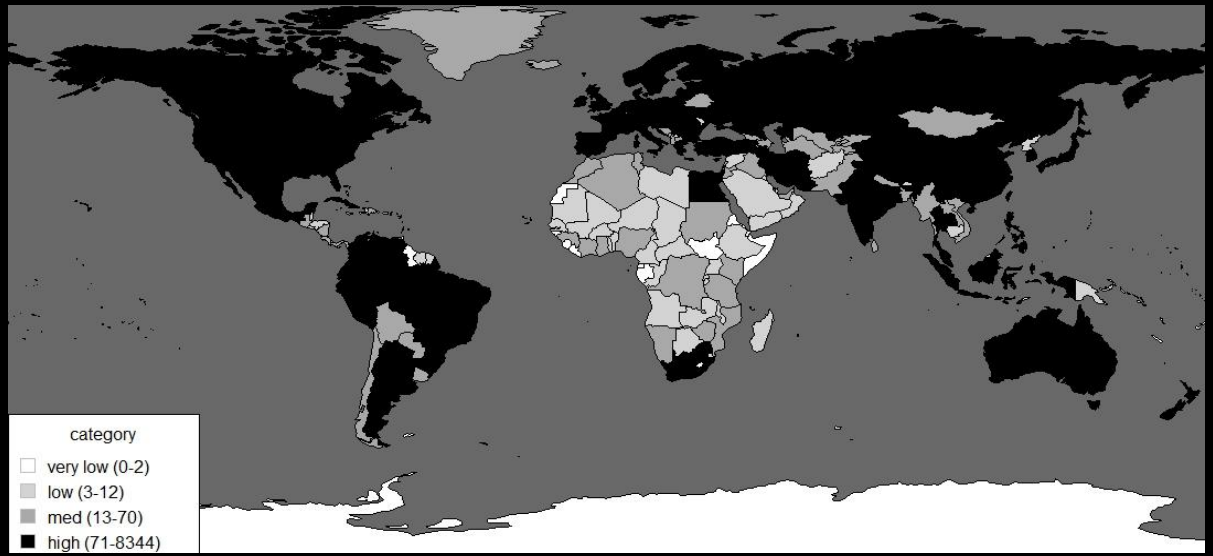
Vision for the future

BY NC SA

The British Museum

museums potential

55,000 museums 180 countries



Alice Daish (2016). RGlobalMuseumMap: First Release.. Zenodo. 10.5281/zenodo.56366

CC BY NC SA

Original Data Source: Museums of the World (2015) Berlin, Boston: De Gruyter Saur. Retrieved 5, Jun 2016, from http://ww.degruyter.com/view/product/448539

The British Museum

Business Problems

=

Data opportunity

Visitor Engagement

Objects

Operations

Buildings

Retail & ecommerce

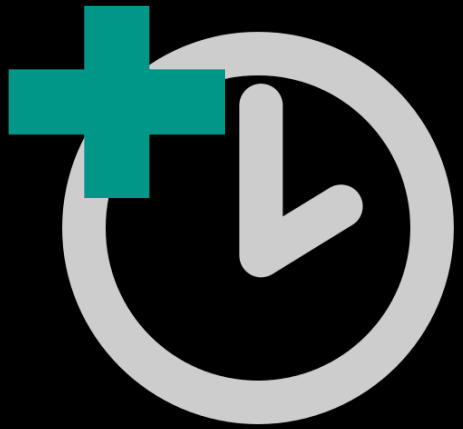
Exhibitions

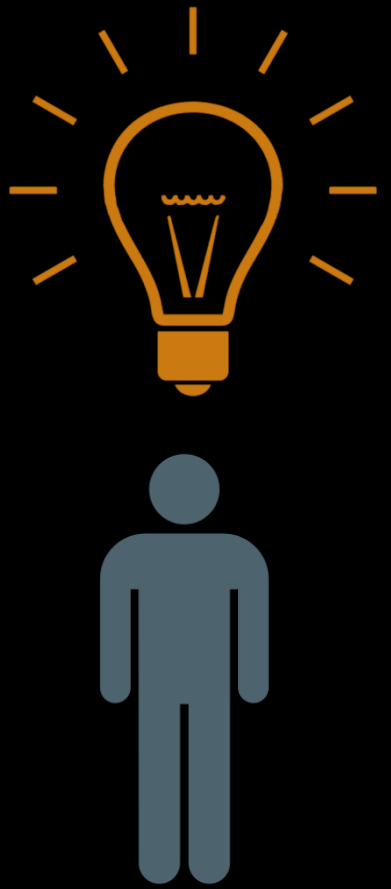
HR

Finance

cc BY NC SA

The British Museum





cc BY NC SA

The British Museum

Thank you & Questions

adaish@britishmuseum.org
@alice_data

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

CC BY NC SA

Symbol - Noun Project: Dara Ullrich, DE