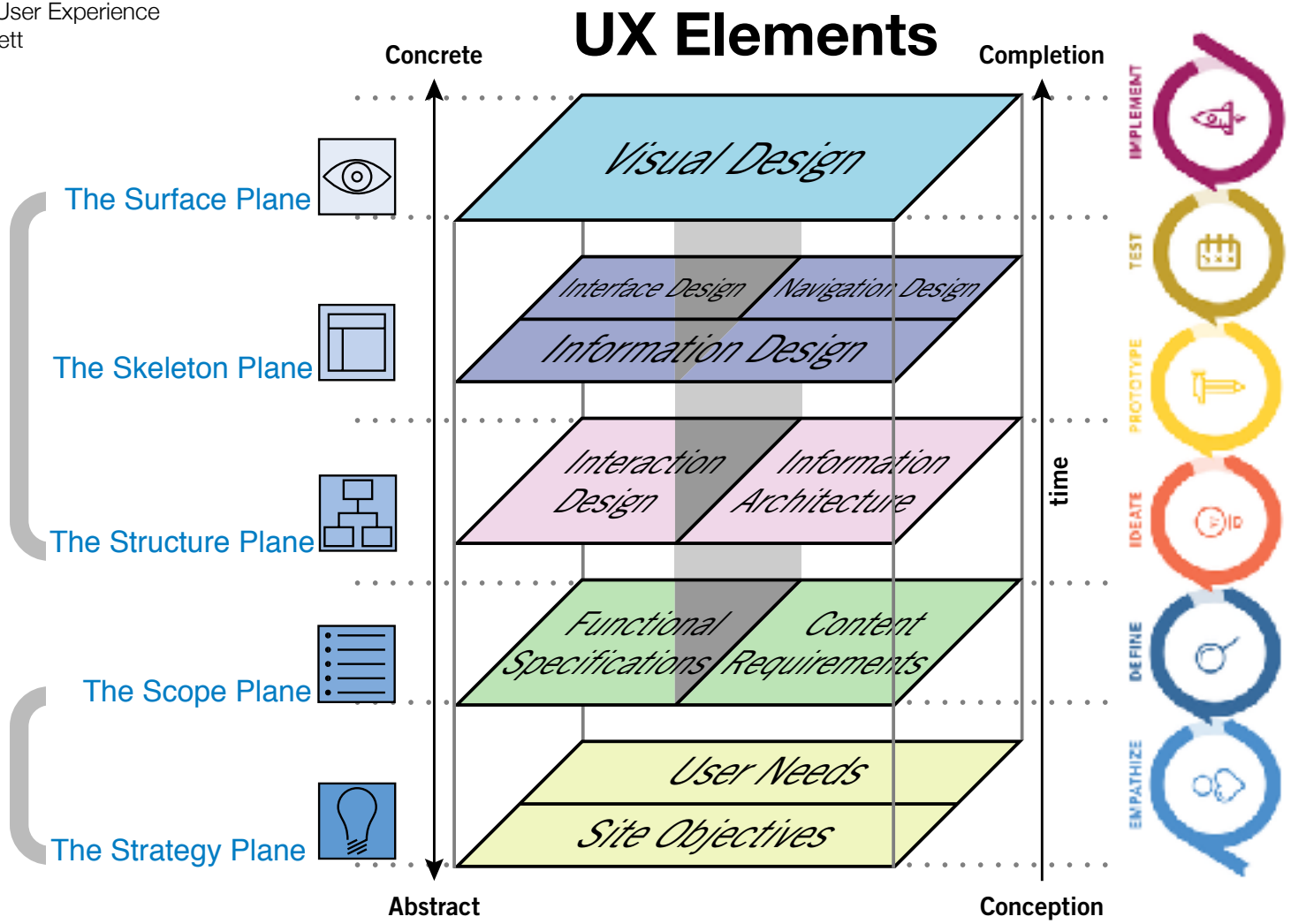
INFORMATION ARCHITECTURE AND DATA MODELLING CYCLE

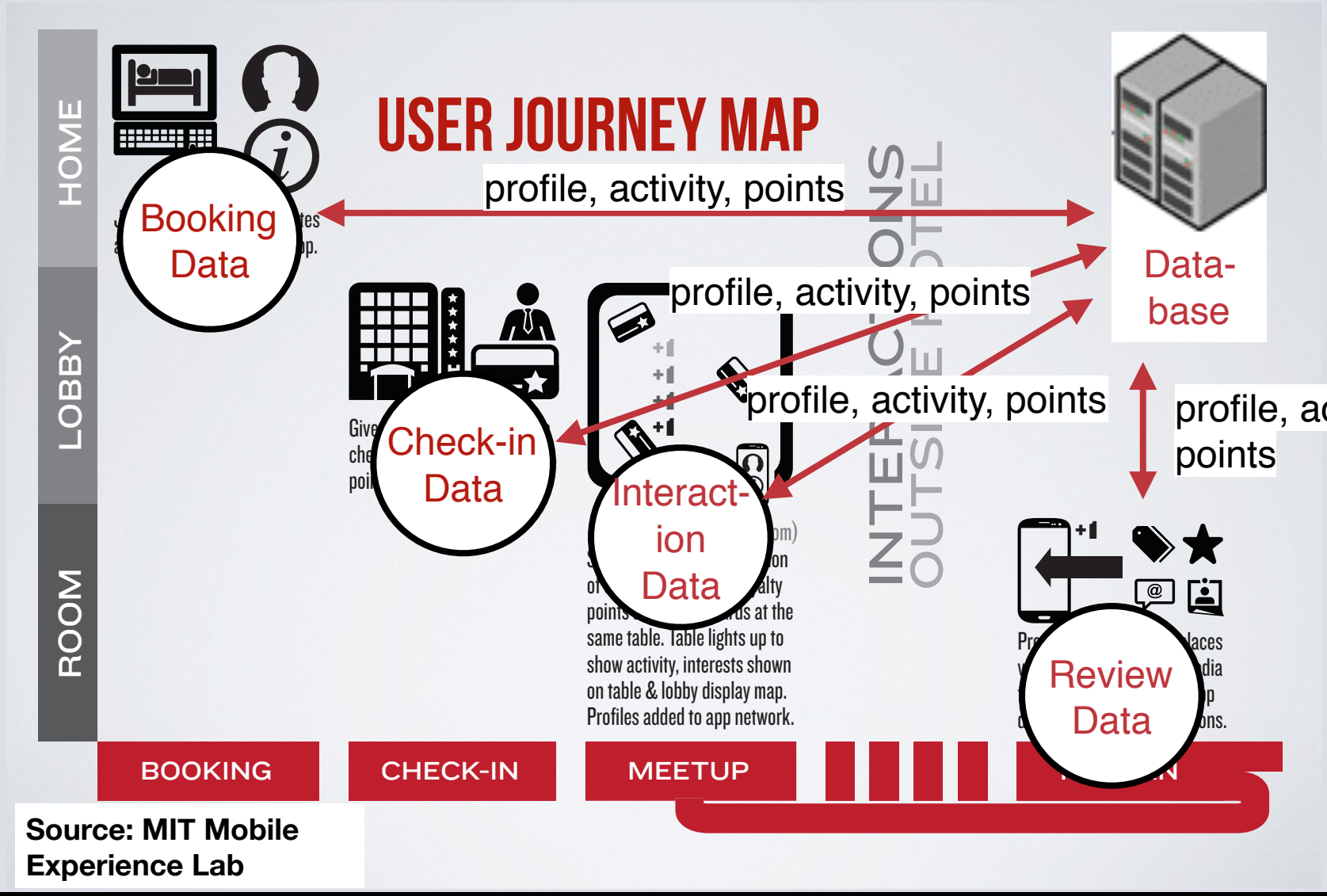
Bernard Suen
Center for Entrepreneurship
Chinese University of Hong Kong

CUHK BUSINESS SCHOOL
Center for Entrepreneurship

DATA MODELLING CYCLE

Source: Elements of User Experience
by Jesse James Garrett

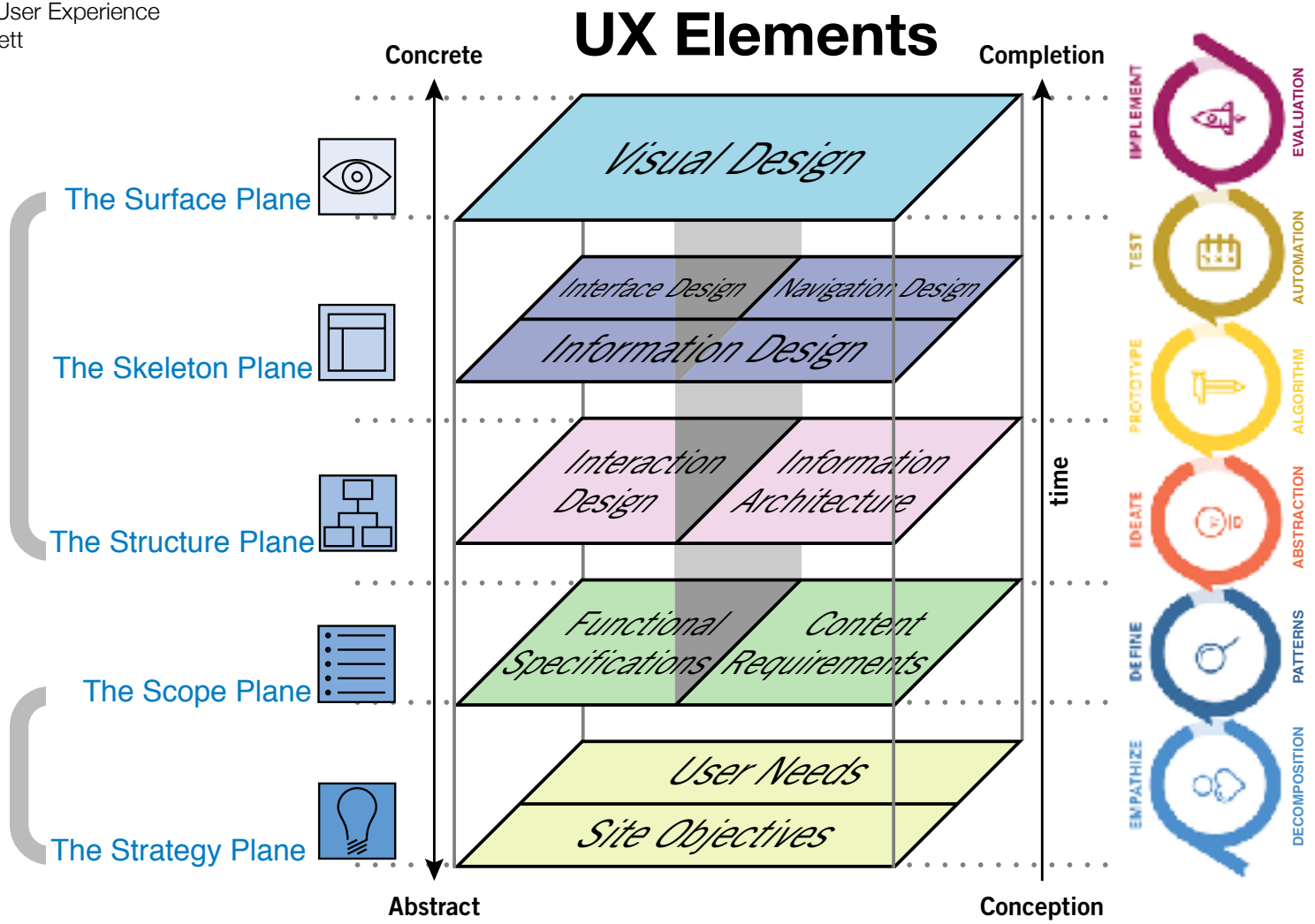




Source: MIT Mobile Experience Lab

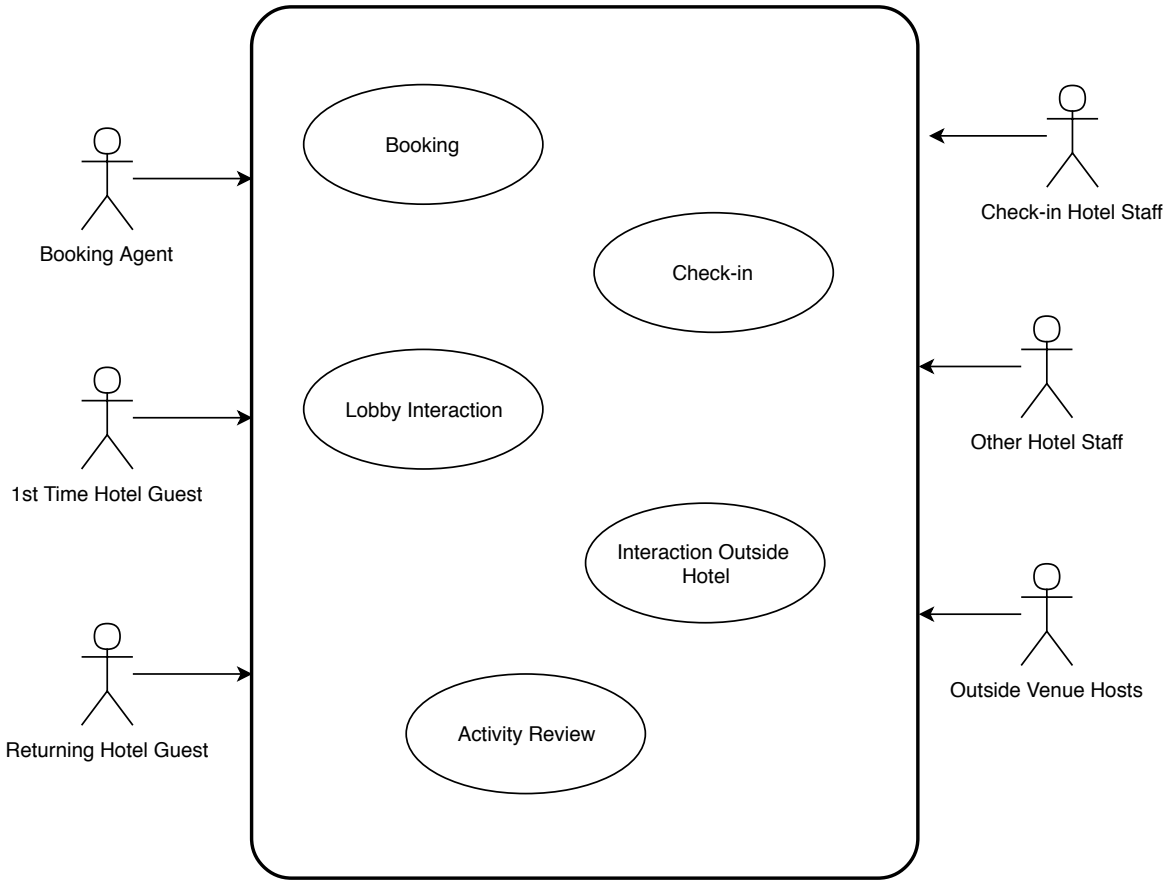
Design Thinking	Computational Thinking	Description
Empathy	Decomposition	Collect and analyse stories and data to understand the stakeholders and discover their needs.
Definition	Patterns	Synthesise recurring persona types, contexts, artefacts, and scenario patterns to formulate problem **(who, what, and why)**.
Ideation	Abstraction	Develop socio-cultural and technical systems to reshape user stories and data flow.
Prototyping	Algorithm	Build experience prototype and computational models to represent future scenarios for validation.
Testing & Implementation	Automation & Evaluation	Continuous testing, improvement and automation to evaluate functional, emotional, social, economic and environmental impacts **(how and how much)**.

Source: Elements of User Experience
by Jesse James Garrett



Develop use cases from journey map.

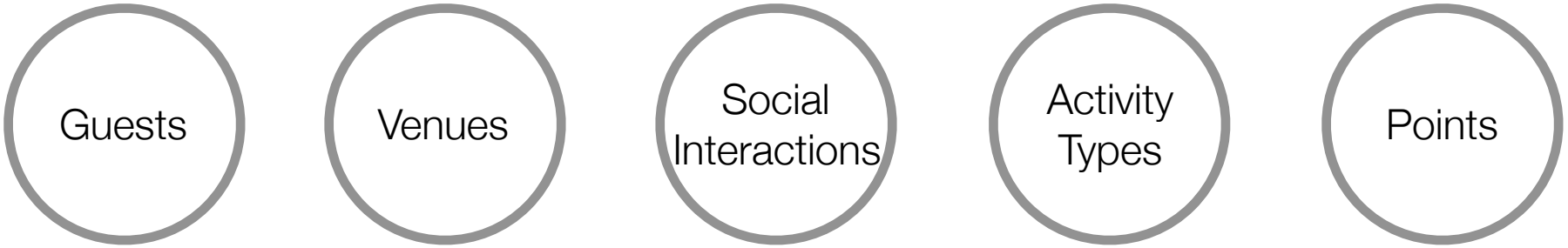
To understand and construct our world through data models.



Use Cases of a Hotel Loyalty App

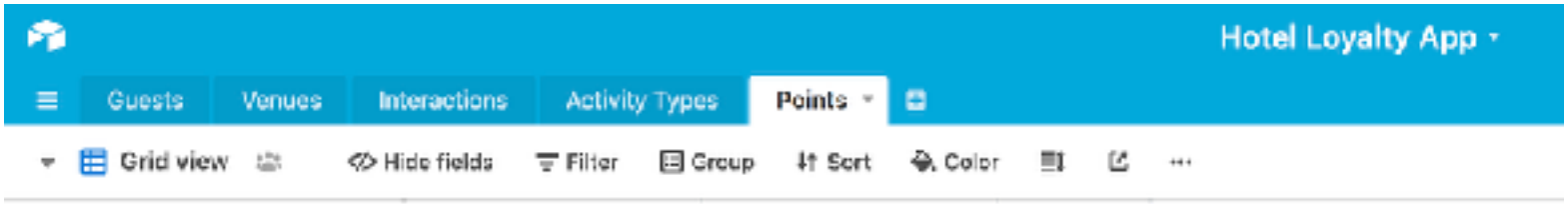
Derive data entities from use cases.

Data Entities to be Tracked



- e.g.
  -1st time guests
  -Returning guests
- e.g.
  -Starbuck
  -Bar inside hotel
  -Boston Metropolitan Museum
- e.g.
  -Add friends
  -Take pictures
  -Share rides
  -Split bills
- e.g.
  -Museum tour
  -Bar hopping
  -Scenic picture taking
  -Business meeting
- e.g.
  -Revisit
  -Dine in hotel
  -Shop in hotel
  -Initiate contact
  -Give reviews

Tracking Data Entities

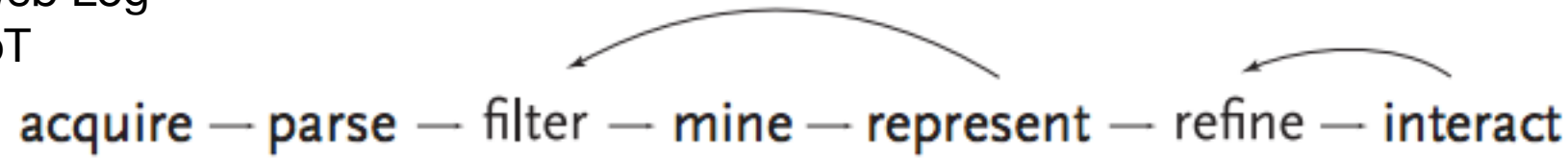


Where does the data come from?

- Open Data
- Scraping (e.g. ParseHub, Scrapy, BeautifulSoup)
- API
- Web Log
- IoT

- Data Science - R
- Python based tools e.g. Pandas, ScikitLearn, TensorFlow
- Knime/Airtable filter, formula and roll-up
- SQL Queries, Data Visualization

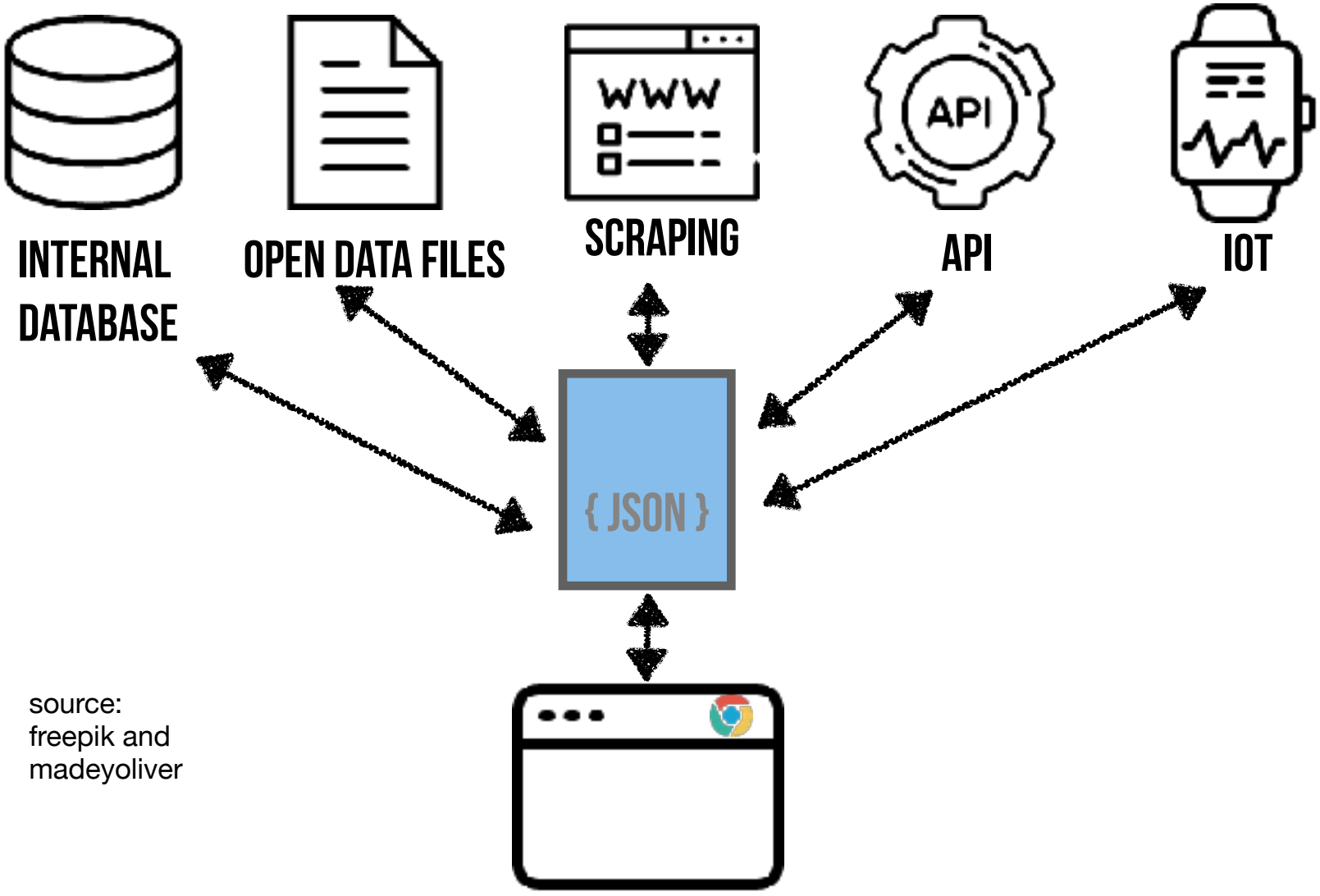
- Charts & Maps (e.g. Tableau)
- JS Mapping/ Charting Tool (e.g. Leaflet, D3)



- Cleaning (e.g. Open Refine)
- Grouping
- Transformation

- Charts & Maps (e.g. Tableau)
- JS Mapping/ Charting Tool (e.g. Leaflet, D3)

source: Ben Fry's Visualizing Data



source:
freepik and
madeyoliver

RISE OF THE SMART CITIES & OPEN DATA

香港智慧城市

Hong Kong Smart City 藍圖 Blueprint

NYC | OpenData 311 | Search all NYC.gov websites

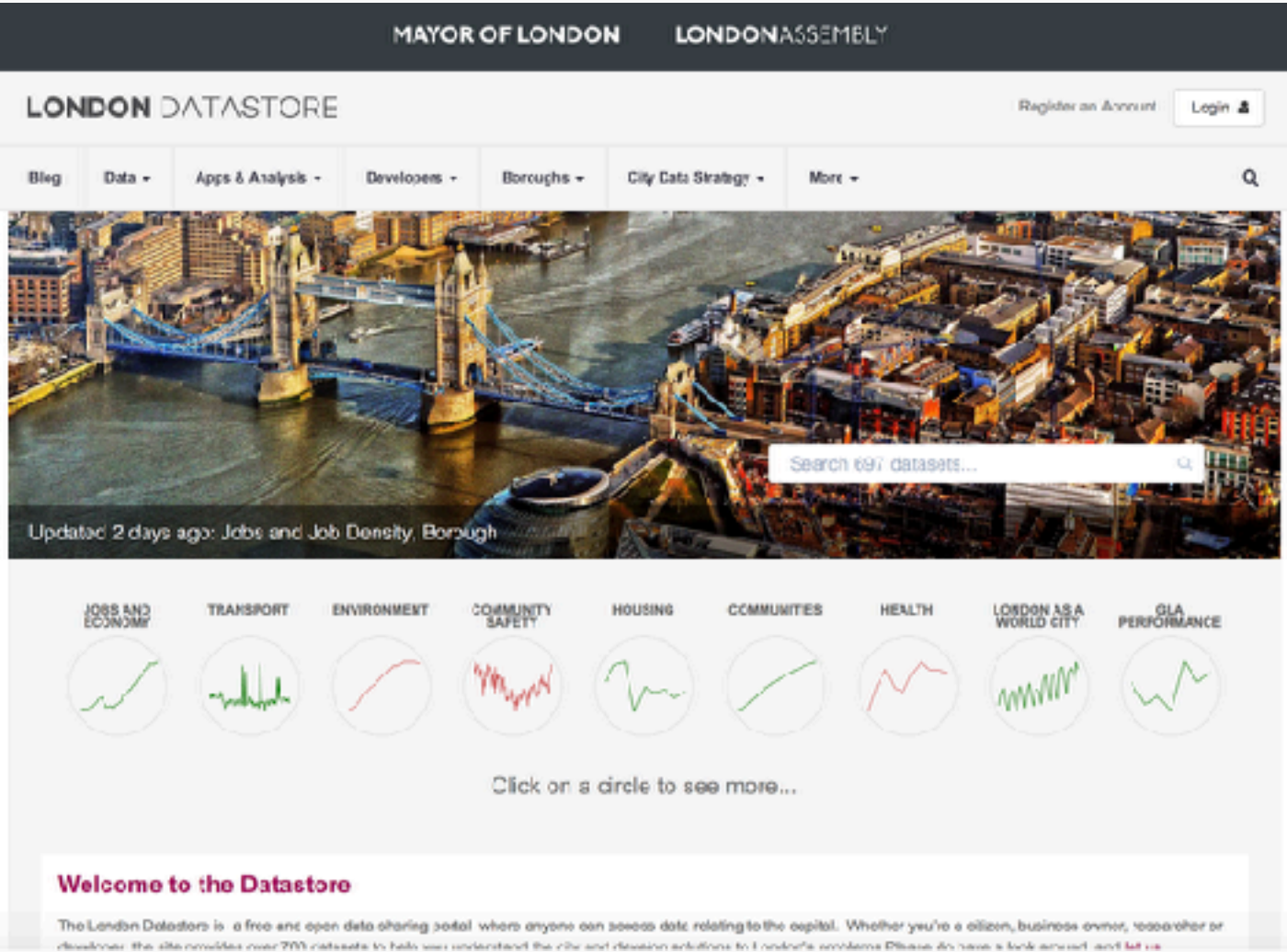
NYC OpenData

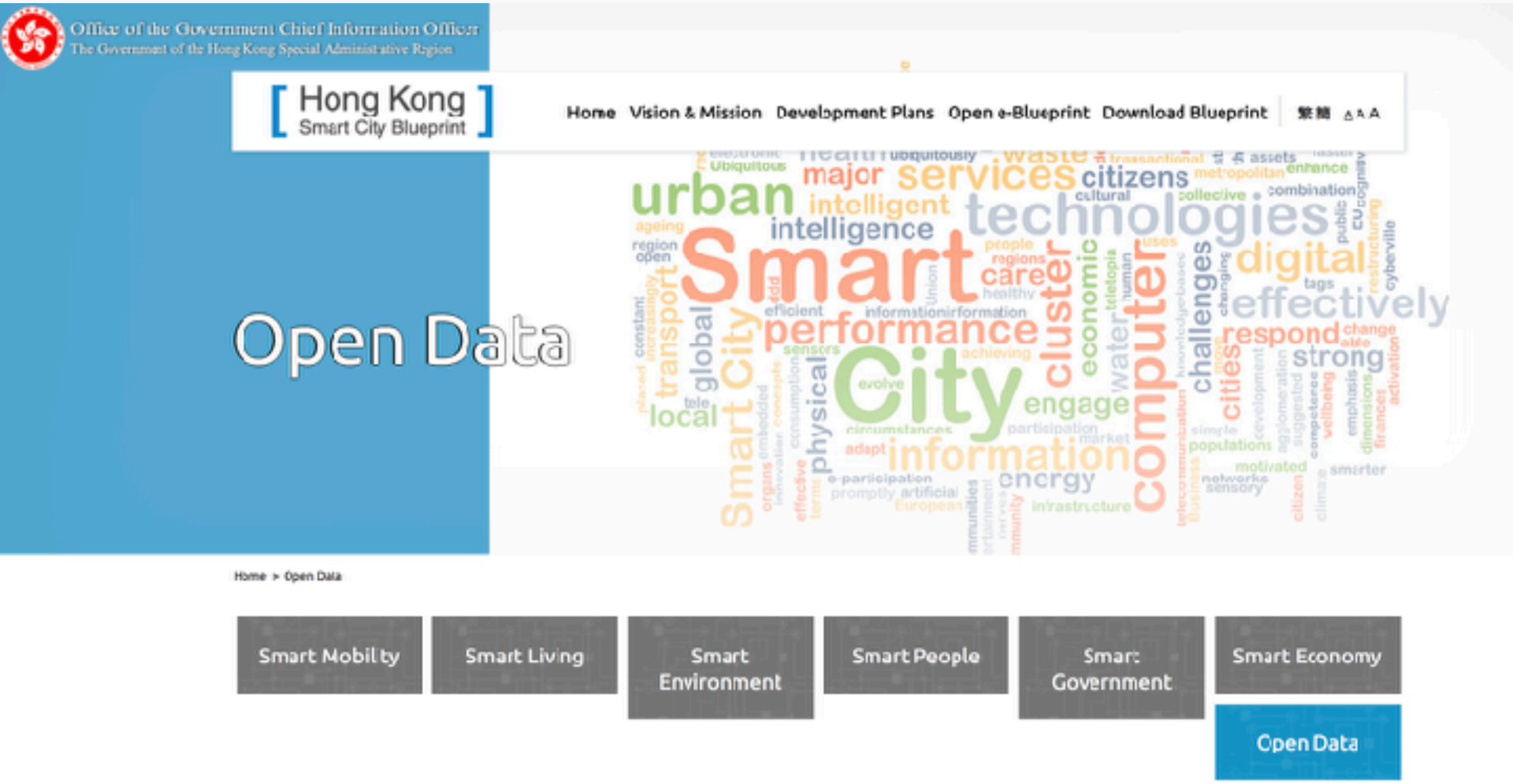
Home Data About ~ Learn ~ Alerts Contact Us Blog

Open Data for All New Yorkers

Where can you find public Wi-Fi in your neighborhood? What kind of tree is in front of your office? Learn about where you live, work, eat, shop and play using NYC Open Data.

Search Open Data for things like 311, Buildings, Crime

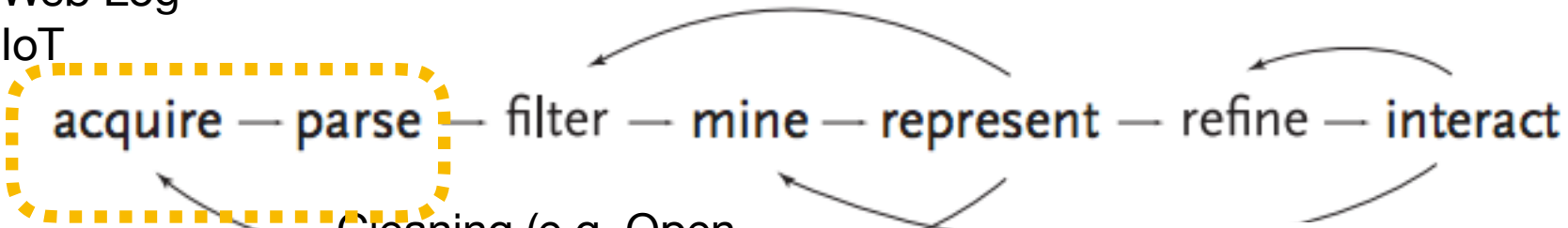

MAYOR OF LONDON
LONDONASSEMBLY
LONDON DATASTORE
Register an Account
Login
Blog
Data
Apps & Analysis
Developers
Boroughs
City Data Strategy
More
Search 697 datasets...
Updated 2 days ago: Jobs and Job Density, Borough
JOBS AND ECONOMY
TRANSPORT
ENVIRONMENT
COMMUNITY SAFETY
HOUSING
COMMUNITIES
HEALTH
LONDON AS A WORLD CITY
GLA PERFORMANCE
Click on a circle to see more...
Welcome to the Datastore
The London Datastore is a free and open data sharing portal where anyone can access data relating to the capital. Whether you're a citizen, business owner, researcher or


Office of the Government Chief Information Officer
The Government of the Hong Kong Special Administrative Region
[Hong Kong
Smart City Blueprint]
Home
Vision & Mission
Development Plans
Open e-Blueprint
Download Blueprint
繁體
Open Data
Home > Open Data
Smart Mobility
Smart Living
Smart Environment
Smart People
Smart Government
Smart Economy
Open Data

- Open Data
- Scraping (e.g. ParseHub, Scrapy, BeautifulSoup)
- API
- Web Log
- IoT

- Data Science - R
- Python based tools e.g. Pandas, ScikitLearn, TensorFlow
- Knime/Airtable filter, formula and roll-up
- SQL Queries, Data Visualization

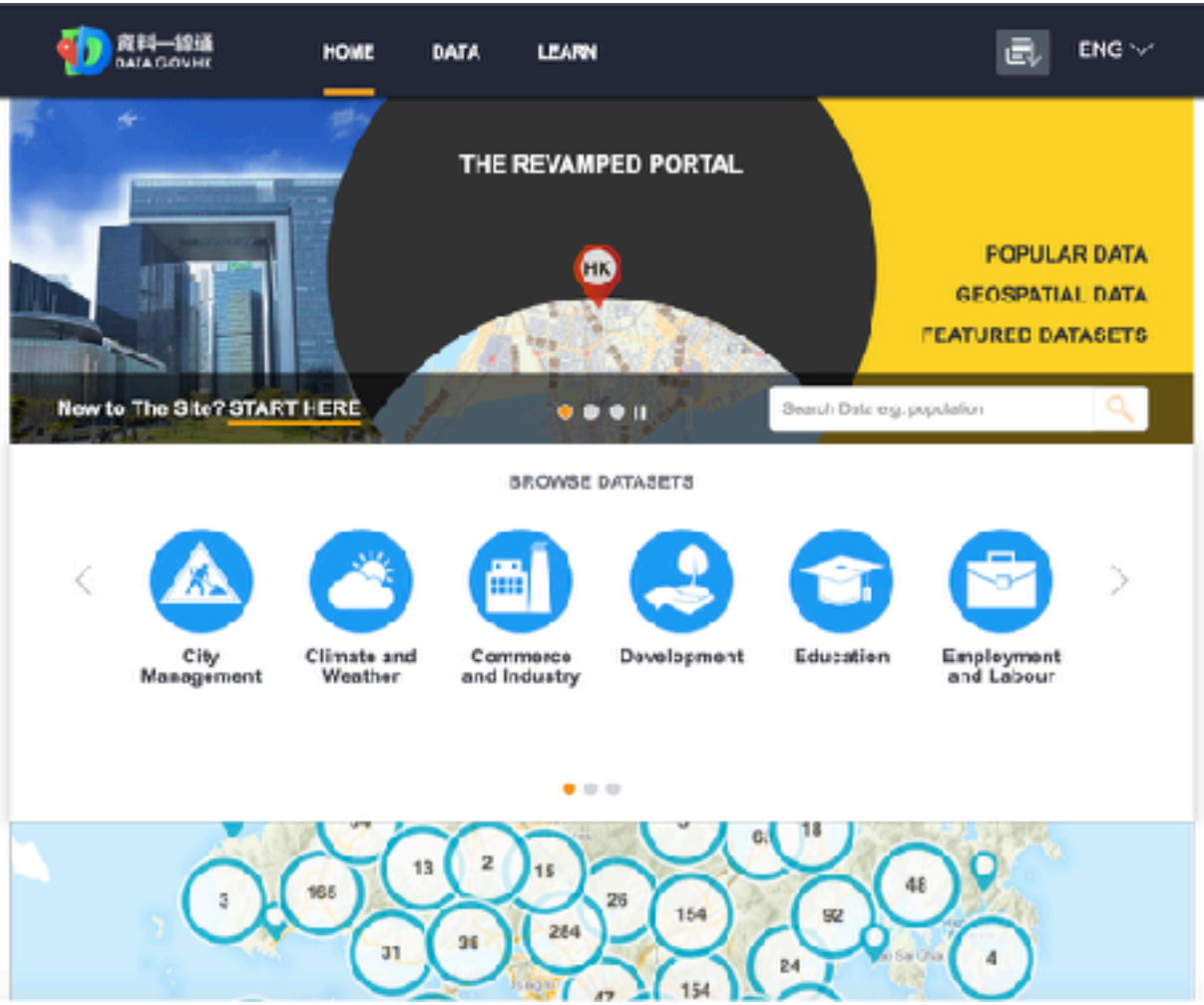
- Charts & Maps (e.g. Tableau)
- JS Mapping/ Charting Tool (e.g. Leaflet, D3)



- Cleaning (e.g. Open Refine)
- Grouping
- Transformation

- Charts & Maps (e.g. Tableau)
- JS Mapping/ Charting Tool (e.g. Leaflet, D3)

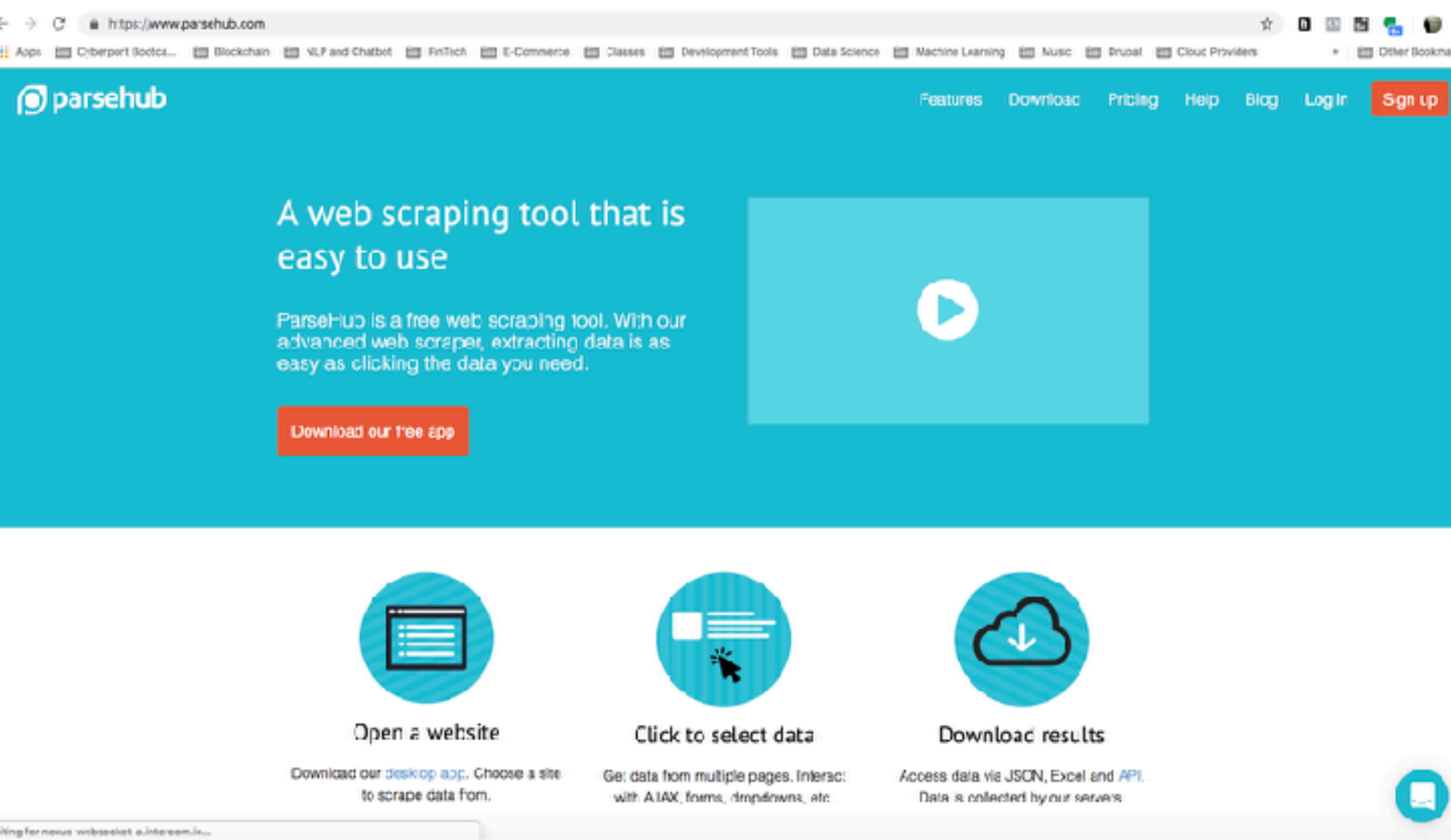
source: Ben Fry's Visualizing Data

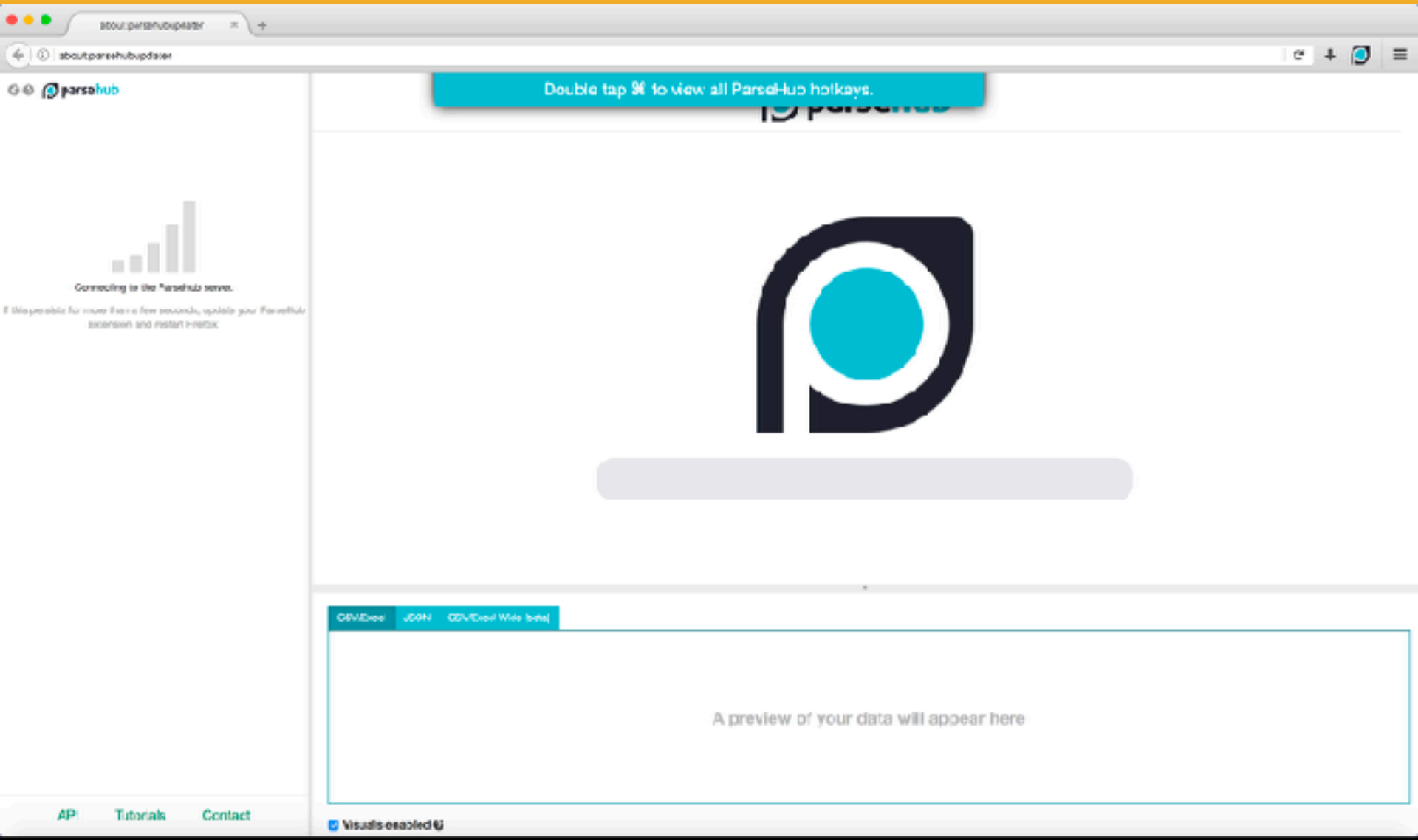

HOME
DATA
LEARN
ENG
THE REVAMPED PORTAL
HK
POPULAR DATA
GEOSPATIAL DATA
FEATURED DATASETS
New to The Site? START HERE
BROWSE DATASETS
City Management
Climate and Weather
Commerce and Industry
Development
Education
Employment and Labour

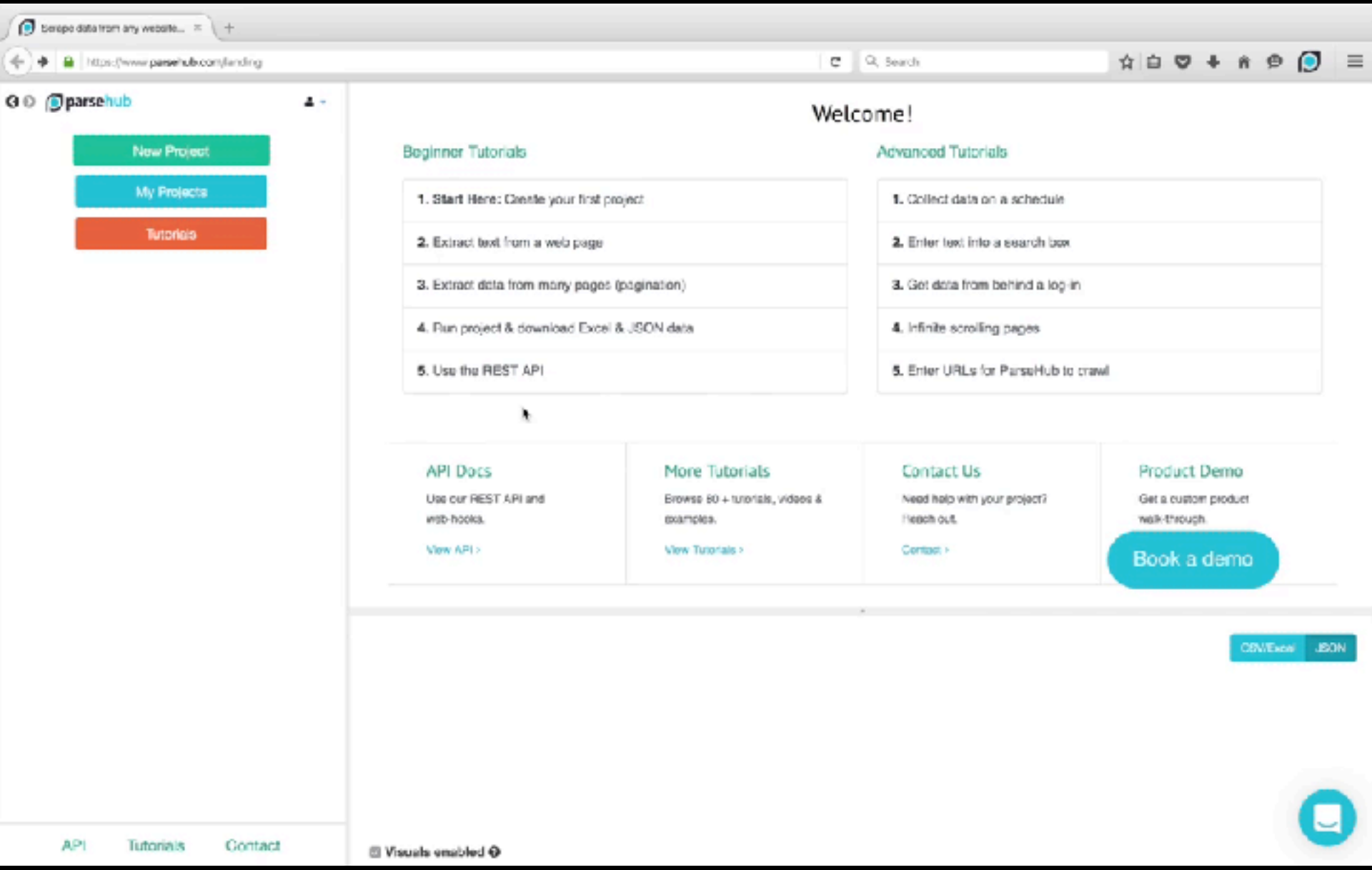
CLASS EXERCISES

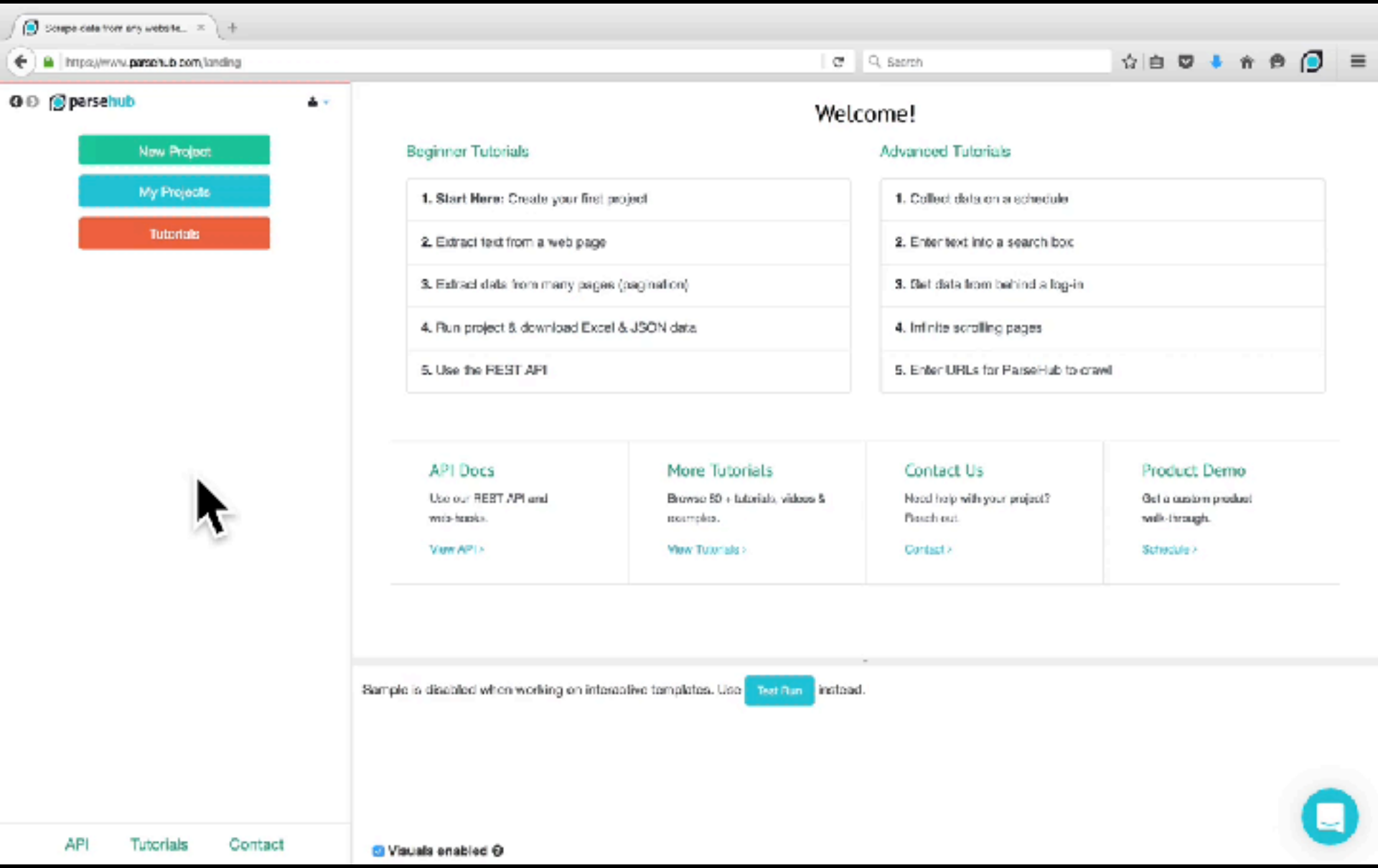
https://www.luxonsoftware.com/converter/xmltocsv

https://json-csv.com/


https://www.parsehub.com
parsehub
Features
Download
Pricing
Help
Blog
Log in
Sign up
A web scraping tool that is easy to use
ParseHub is a free web scraping tool. With our advanced web scraper, extracting data is as easy as clicking the data you need.
Download our free app
Open a website
Download our desktop app. Choose a site to scrape data from.
Click to select data
Get data from multiple pages. Interact with AJAX, forms, dropdowns, etc.
Download results
Access data via JSON, Excel and API.
Data is collected by our servers


about:parsehubupdater
about:parsehubupdater
parsehub
Double tap ⌘ to view all ParseHub hotkeys.
Connecting to the ParseHub server.
A preview of your data will appear here
API
Tutorials
Contact


parsehub
New Project
My Projects
Tutorials
Welcome!
Beginner Tutorials
1. Start Here: Create your first project
2. Extract text from a web page
3. Extract data from many pages (pagination)
4. Run project & download Excel & JSON data
5. Use the REST API
Advanced Tutorials
1. Collect data on a schedule
2. Enter text into a search box
3. Get data from behind a log in
4. Infinite scrolling pages
5. Enter URLs for ParseHub to crawl
API Docs
Use our REST API and web hooks.
View API >
More Tutorials
Browse 60 + tutorials, videos & examples.
View Tutorials >
Contact Us
Need help with your project? Reach out.
Contact >
Product Demo
Get a custom product walk-through.
Book a demo
CSV/Excel
JSON
API
Tutorials
Contact
Visuals enabled


parsehub
New Project
My Projects
Tutorials
Welcome!
Beginner Tutorials
1. Start Here: Create your first project
2. Extract text from a web page
3. Extract data from many pages (pagination)
4. Run project & download Excel & JSON data
5. Use the REST API
Advanced Tutorials
1. Collect data on a schedule
2. Enter text into a search box
3. Get data from behind a log-in
4. Infinite scrolling pages
5. Enter URLs for ParseHub to crawl
API Docs
More Tutorials
Contact Us
Product Demo
Sample is disabled when working on interactive templates. Use Test Run instead.
API
Tutorials
Contact
Visuals enabled

http://www.dianping.com/hongkong/food

Not Secure | openrefine.org

Apps | Cyberport Bootca... | Blockchain | NLP and Chatbot | FinTech | E-Commerce | Classes | Development Tools | Data Science | Machine Learning | Music | Drupal | Cloud Providers | Other Bookmarks

OpenRefine

A free, open source, powerful tool for working with messy data

- Home
- Community
- Documentation
- Download
- Contact Us
- Blog

Enhanced with Java profiler

JPROFILER

Welcome!

OpenRefine (formerly Google Refine) is a powerful tool for working with messy data: cleaning it; transforming it from one format into another; and extending it with web services and external data.

OpenRefine is available in English, Chinese, Spanish, French, Russian, Portuguese (Brazil), German, Japanese, Italian, Hungarian, Hebrew, Filipino, Cebuano, Tagalog

OpenRefine is supported by:

Google News Initiative

Introduction to OpenRefine

1. Explore Data

OpenRefine can help you explore large data sets with ease. You can find out more about this functionality by watching the video below and going through these articles

Google refine

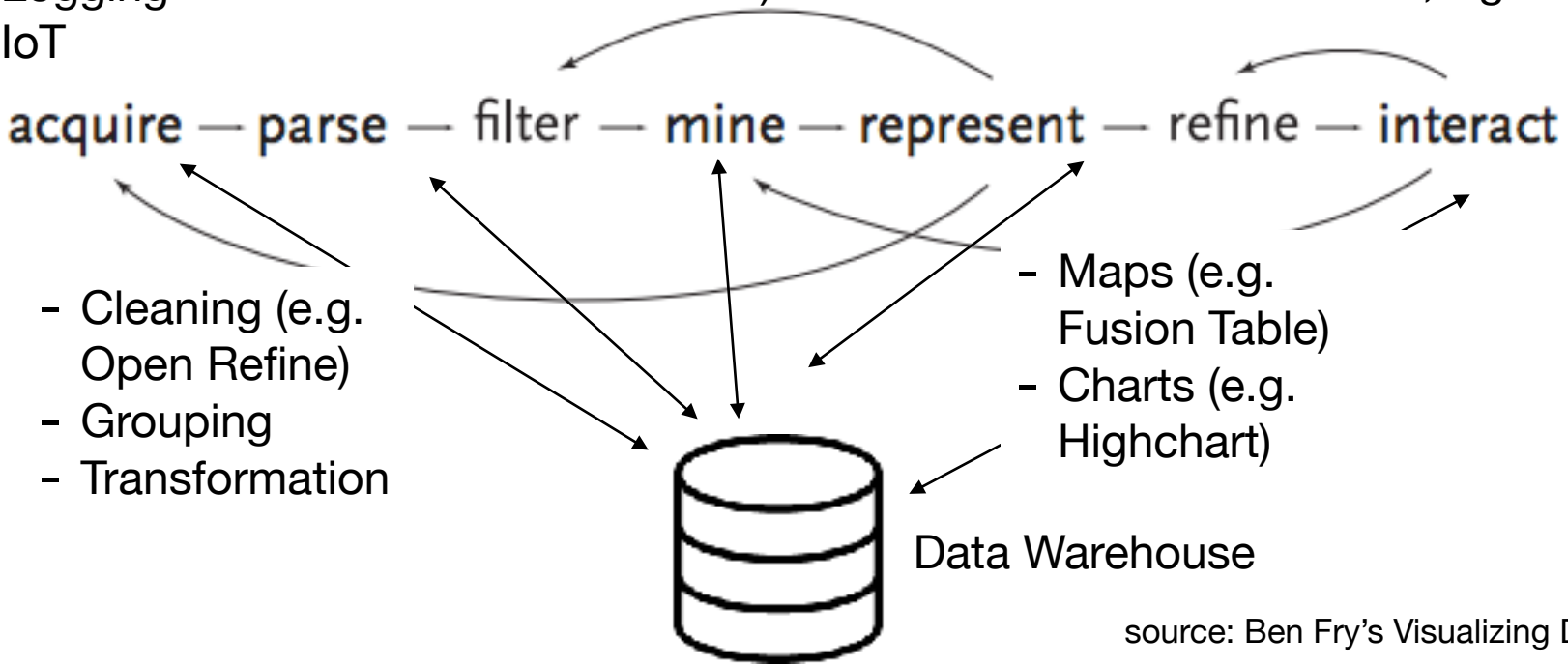
http://d3-media.blogspot.hk/2013/11/how-to-refine-your-data.html

HOW DOES THE DATA VISUALIZATION CYCLE WORK WITH DATABASE MANAGEMENT?

- Open Data
- Scraping (e.g. Portia)
- API
- Logging
- IoT

- Statistics
- AI (ML/DL)
- SQL Queries (e.g. Where/Count/Avg/ Max/Min in Drupal Views)

- Maps
- Charts (e.g. D3, ChartJS,Highchart)



- Cleaning (e.g. Open Refine)
- Grouping
- Transformation

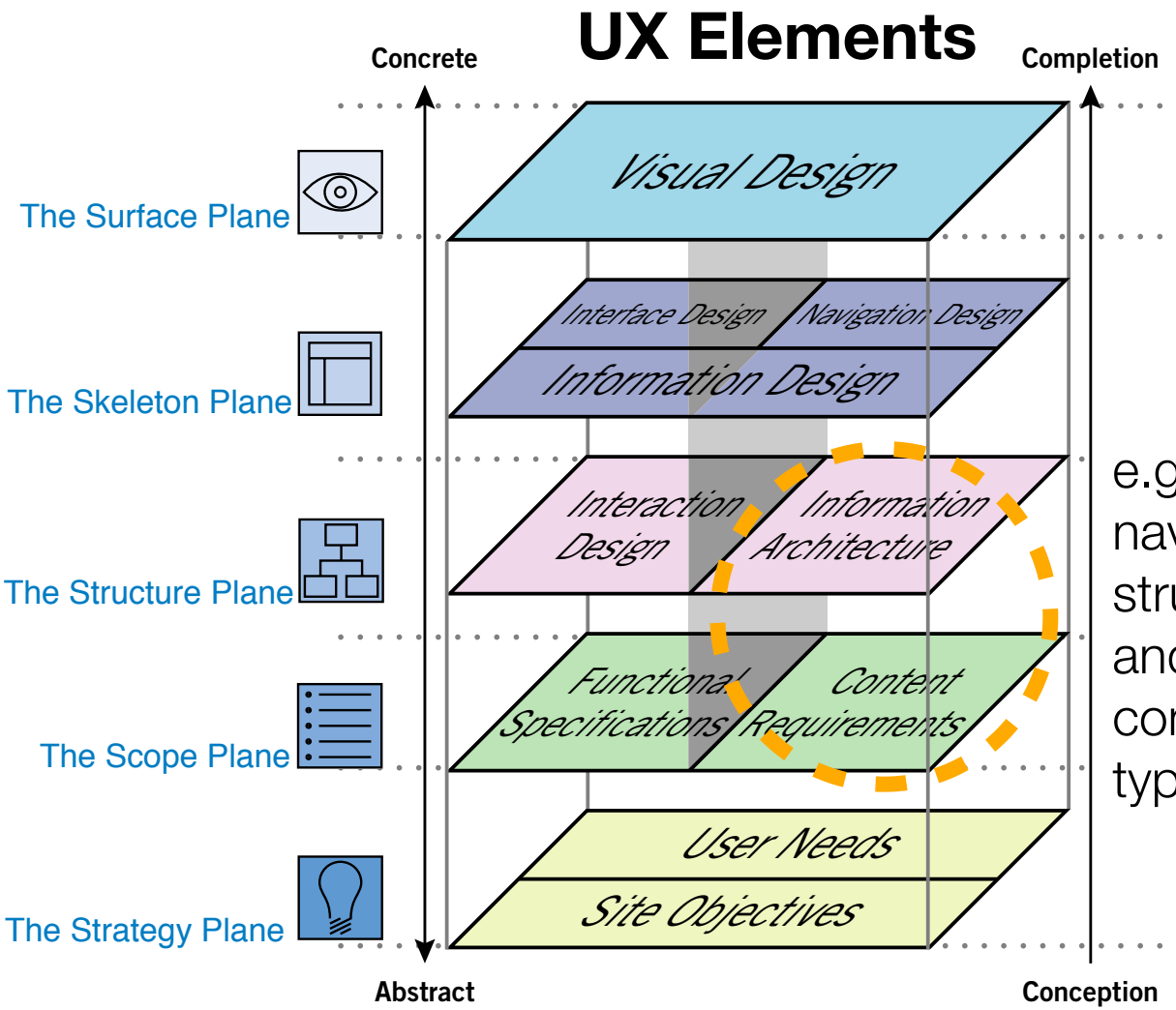
- Maps (e.g. Fusion Table)
- Charts (e.g. Highchart)

source: Ben Fry's Visualizing Data

FROM DATABASE MANAGEMENT TO CONTENT MANAGEMENT AND INFORMATION ARCHITECTURE

Can you see a visualization process here?

Where does data fit?



e.g. navigation structure and content types

Source: Elements of User Experience by Jesse James Garrett

Empathise
身同感受

Define
界定问题

Ideate
創意發想

Prototype
開展原型

Test &
Implement
測試執行

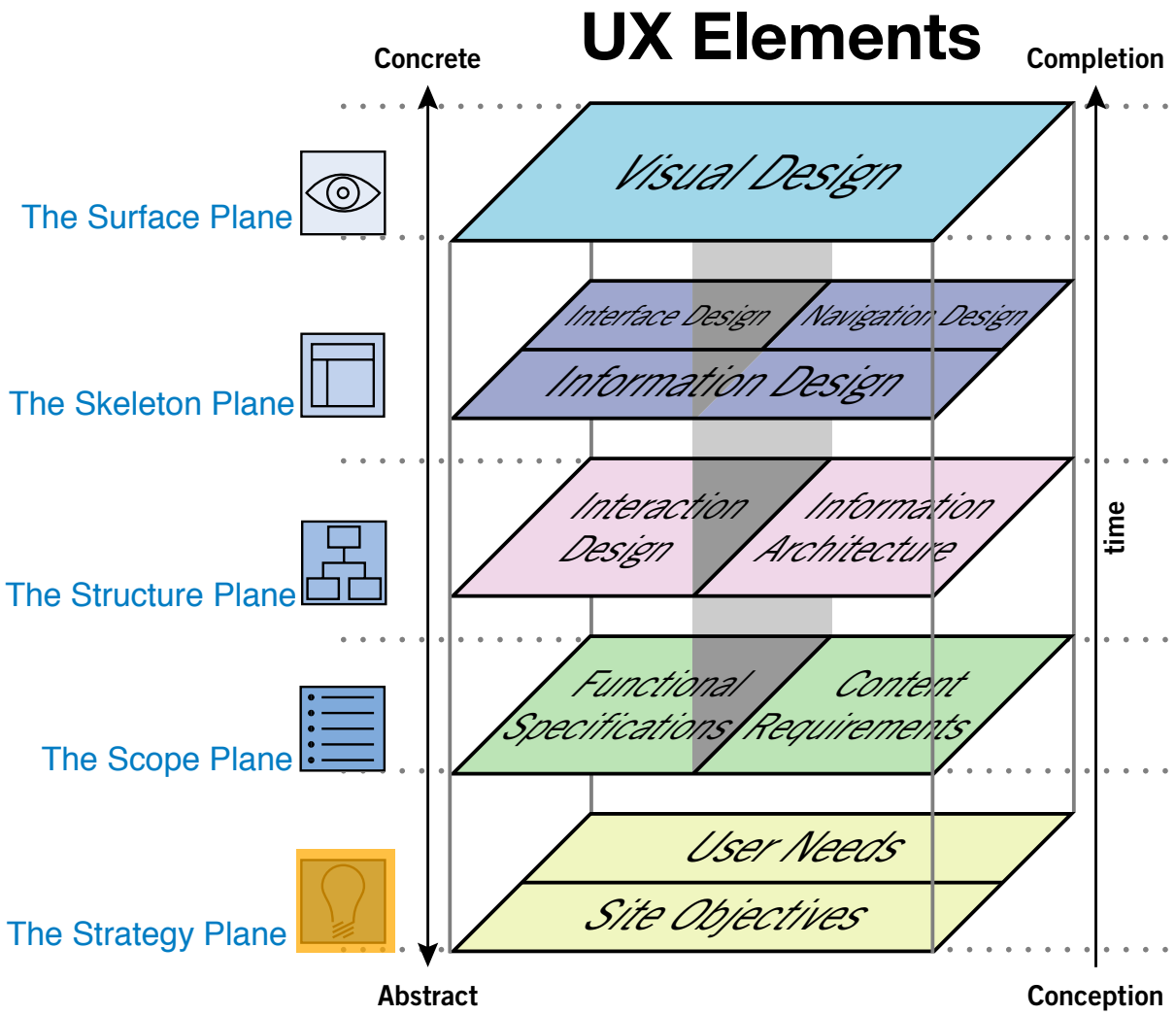
THE 5 STEP DESIGN THINKING JOURNEY

Source: Stanford D.School

EXPERIENCE = JOURNEY

e.g. The Journey of a New CUHK Applicant

EXAMPLE:
THE JOURNEY OF A NEW CUHK APPLICANT



Source: Elements of User Experience
by Jesse James Garrett

香港中文大學
The Chinese University of Hong Kong

入學及學生資助處
Office of Admissions and Financial Aid

About CUHK

Welcome Message

I am delighted that you are considering The Chinese University of Hong Kong for your undergraduate studies.

Founded in 1963 with a mission "to combine tradition with modernity, and to bring together China and the West", the University has, since its founding, developed into an academic institution of international standing, well-known for its excellence in both teaching and research.

Empathise
身同感受

Define
界定问题

Ideate
創意發想

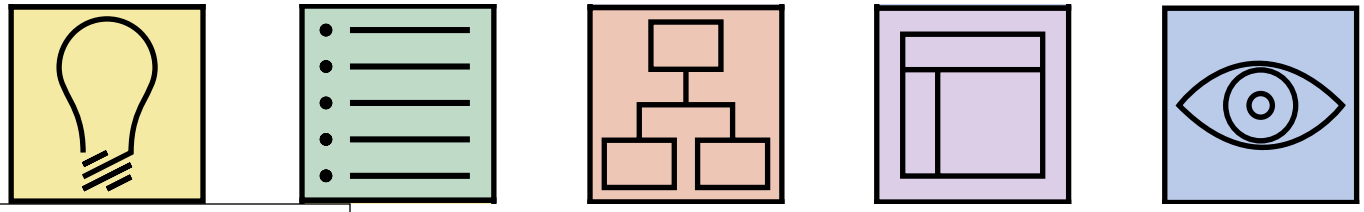
Prototype
開展原型

Test &
Implement
測試執行

THE 5 STEP DESIGN THINKING JOURNEY

Source: Stanford D.School

SITE OBJECTIVES		TARGET USERS		USER NEEDS
Primary goal	Become top tier university in the world.	Primary User	Top local students and world-class academics	Apply for degree enrollment and job vacancies.
Additional goal	Increase research output.	Secondary User	Chinese students interested in the west	Interested to come to CUHK to study.
Additional goal	Build strong alumni network.	Secondary User	Foreign students interested in PRC	Will consider CUHK for full-degree and exchange program.



STRATEGY PLANE

Empathise
身同感受

Define
界定问题

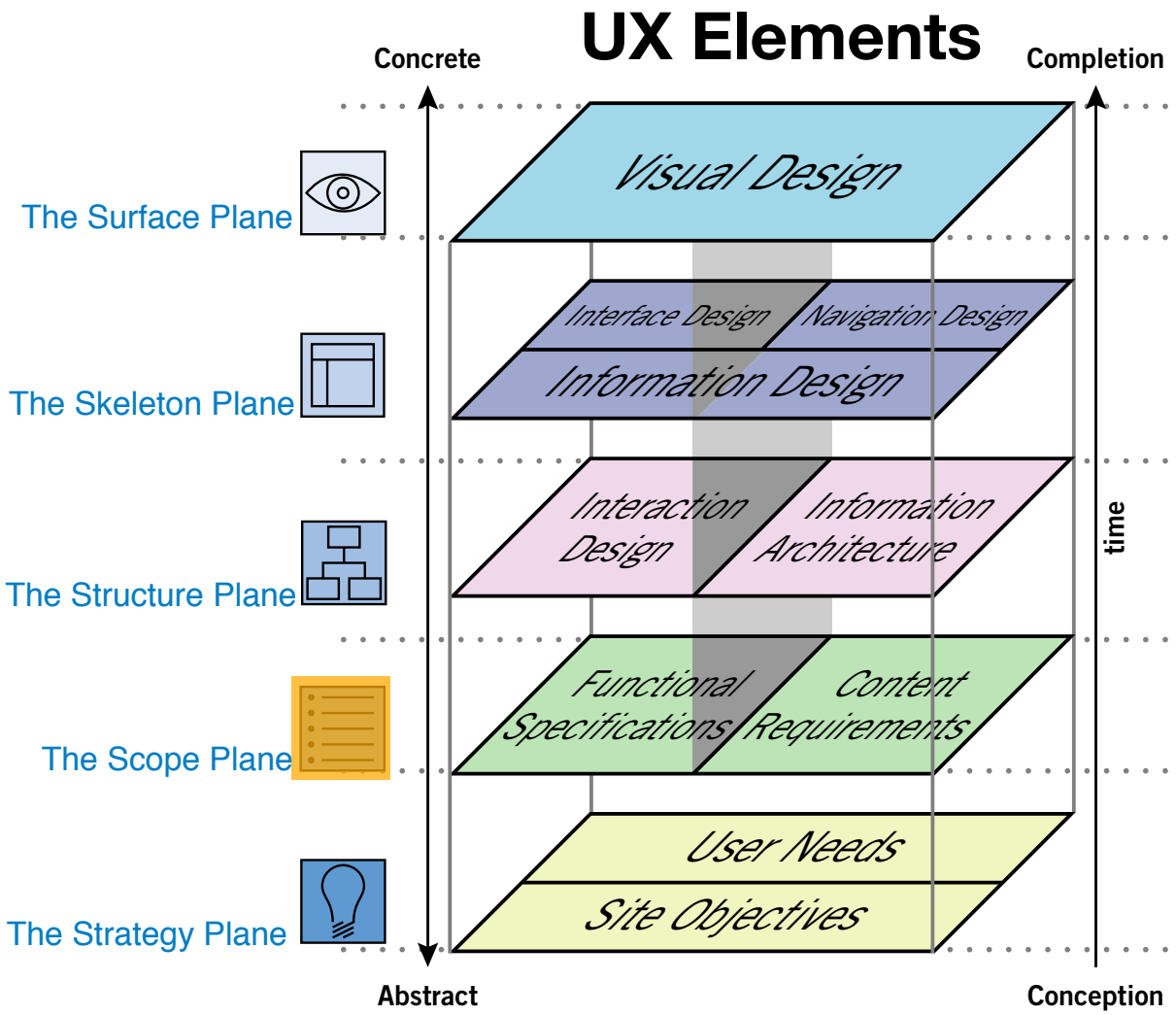
Ideate
創意發想

Prototype
開展原型

Test &
Implement
測試執行

THE 5 STEP DESIGN THINKING JOURNEY

Source: Stanford D.School



Source: Elements of User Experience
by Jesse James Garrett

THE JOURNEY AS A STORY

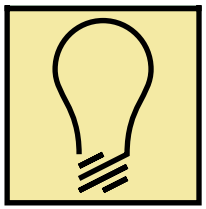
ELEMENTS

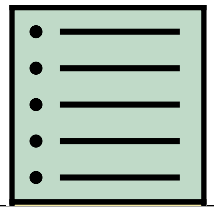
1. 人 (人物) PERSONA
2. 景 (場景) CONTEXT
3. 物 (物件) ARTIFACTS
4. 用 (使用) SCENARIOS

STRUCTURE

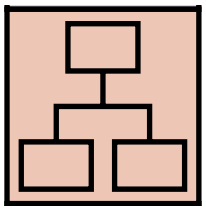
1. 起 (人物背景) BEGINNING
2. 承 (問題) CHALLENGE
3. 轉 (發展方案) RESPONSE
4. 合 (完成狀況) END

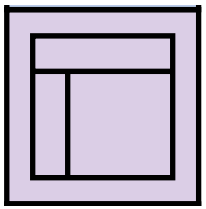
資料來源---劇本導引:
資訊時代產品與服務設計新法
作者余德彰, 林文綺, 王介丘

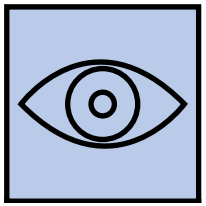




SCOPE PLANE







Journey Map (歷程圖)

功用: 有助收集和分類用戶的故事。

參考資料來源: Designthinker Academy

有助描繪用戶、器物或系統的故事，以人景物用的故事元素、起承轉合的故事結構，更好地理解目標對象的需求，如何在故事中浮現，並尋找設計機會。例如用戶日誌，用戶的生活地圖，或產品如何在空間中移動的流程（從製造商到商店貨架，再轉送到用戶的手上）。

	起	承	轉	合
人物				
場景				
文物				
情況				

	起	承	轉	合
人物	- Local - PRC/ International	Curious about the school and program	Getting serious and anxious	Excitement
場景	online promotion	school recruitment	CUHK	online/email
文物	website	brochure/ presentation	- campus map - hk map - application form	letter of acceptance
情況	online research Persona wants to filter program	staff roadshow Persona meets professor and CUHK staff	school visit Check out the school	application Finalize options	admission Accept offer

例子

TYPES OF PERSONA (e.g. STUDENT & STAFF)

Persona (人物輪廓)

功用: 有助分類不同人物的輪廓。

參考資料來源: Interaction Design Foundation

人物輪廓包括個人檔案（年齡、性別、教育程度、家庭關係、職業、收入等），生活方式偏好，購物和媒體使用習慣，人生目標和生活日程。工作小組必須為人物輪廓的類型數目達成共識，並且為每種類型命名，方便大家溝通。Persona 也可算在設計思維中很重要的pattern. 不同的persona有不同需要。

 名字: 年齡: 性別: 教育程度: 家庭關係: 職業:	生活方式偏好: 購物習慣: 媒體使用習慣: 人生目標: 生活日程:

OBSERVATION AND UNDERSTANDING

LOCAL APPLICANTS

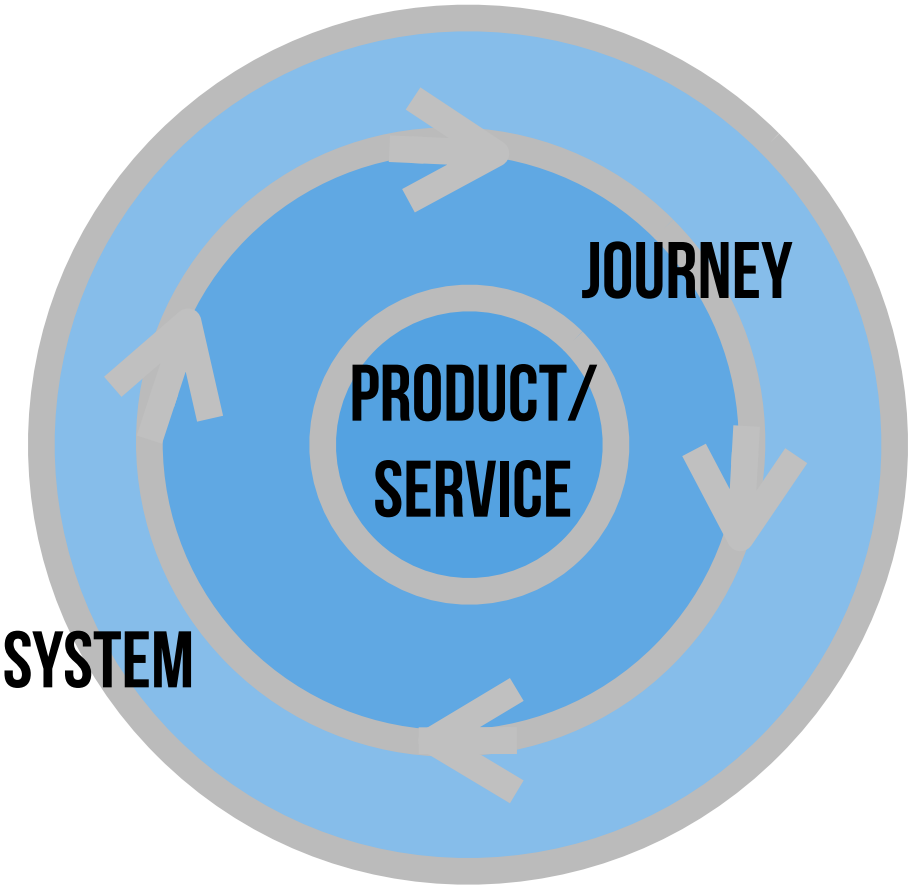
FOREIGN APPLICANTS

MAINLAND APPLICANTS

PROBLEM STATEMENT:

How can CUHK's website help new student and staff applicants pursue their career objectives over the alternatives out there while at the same time communicate the school's brand to continue attracting the best?

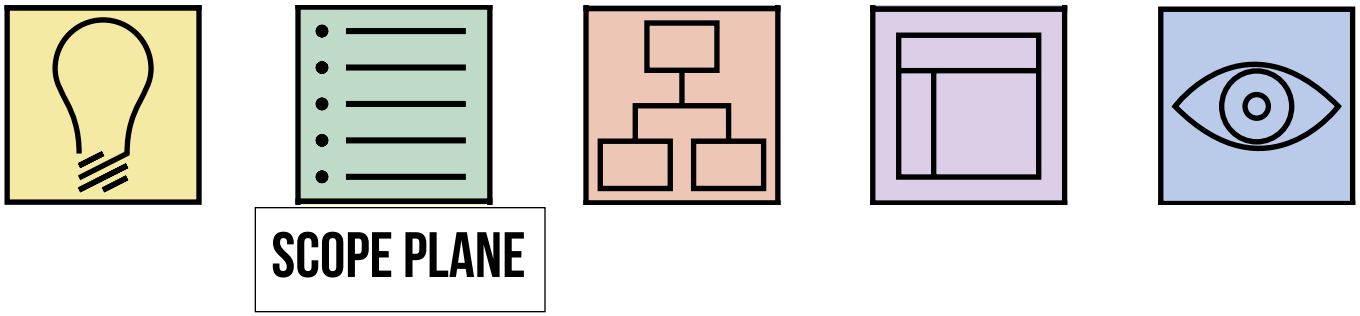
THE BIGGER PICTURE



- Product/Service: the offering (e.g. enrollment at CUHK) which is designed to address a need (e.g. world-class education).
- Journey: the process between a start and end point through which the user can have one's needs satisfied by accepting the offering (e.g. from gathering school info to final admission).
- System: the technical and social systems (e.g. school/faculty administration/research/teaching/student life) existing in an environment where competitive alternatives (e.g. other schools/programs) exist to address the same need. The question is: what are the compelling reasons for the persona to choose one offering over the others?

Content Production Plan

Level	Section	Description	Status	Deadline	Handled by
1.0	About Us	An overview of the organization	Current	NA	Peter
1.1	History	Historical background & timeline	In progress	15/10/11	Mary
1.2	Contact Us	Telephone no., address, and contact	New	22/10/11	John
2.0	Product	Listing of current products and services	In progress	25/10/11	Henry



Empathise
身同感受

Define
界定问题

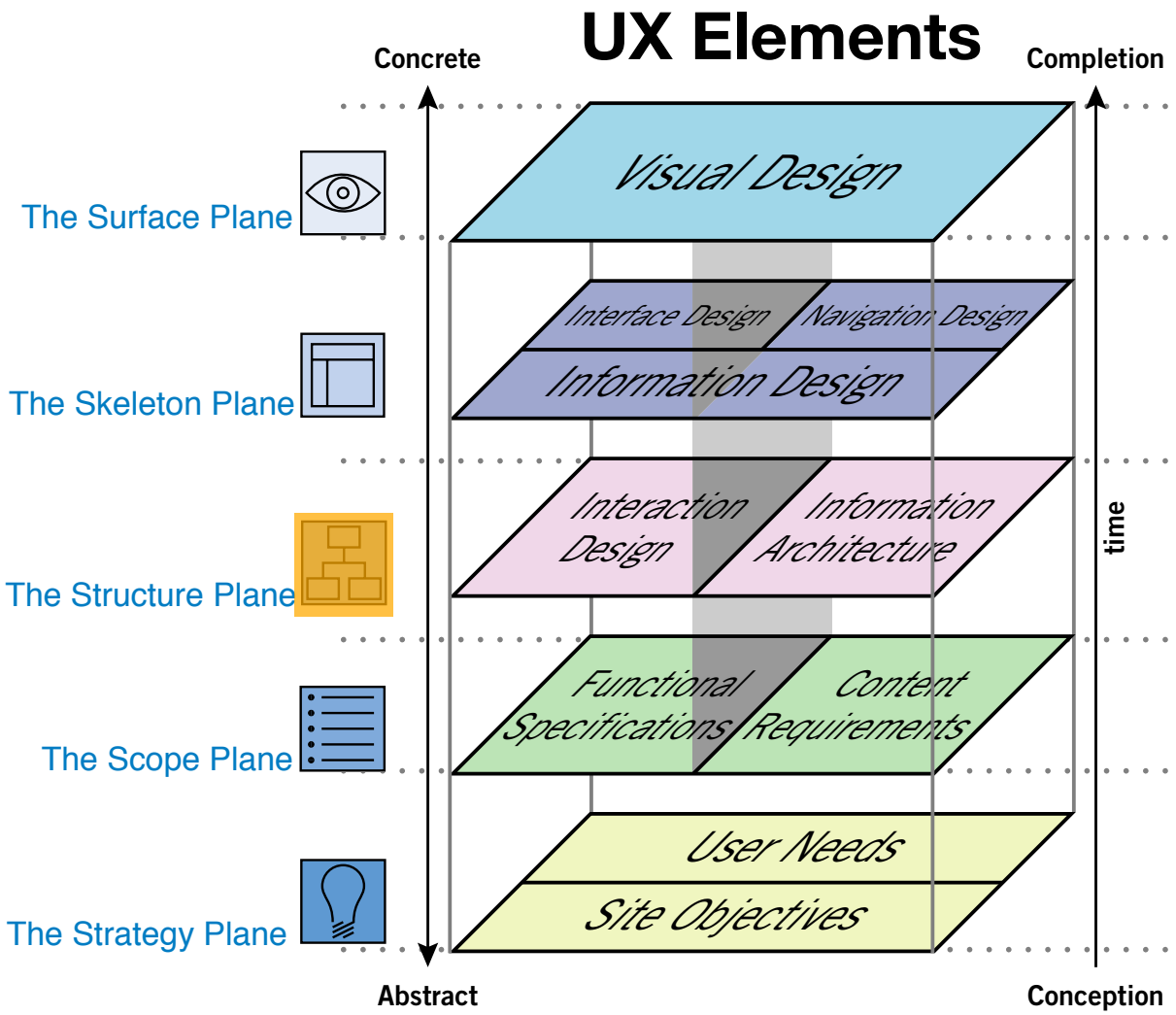
Ideate
創意發想

Prototype
開展原型

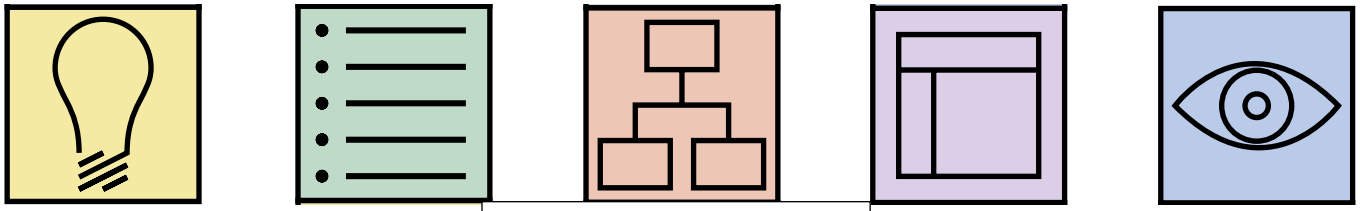
Test &
Implement
測試執行

THE 5 STEP DESIGN THINKING JOURNEY

Source: Stanford D.School



Source: Elements of User Experience
by Jesse James Garrett

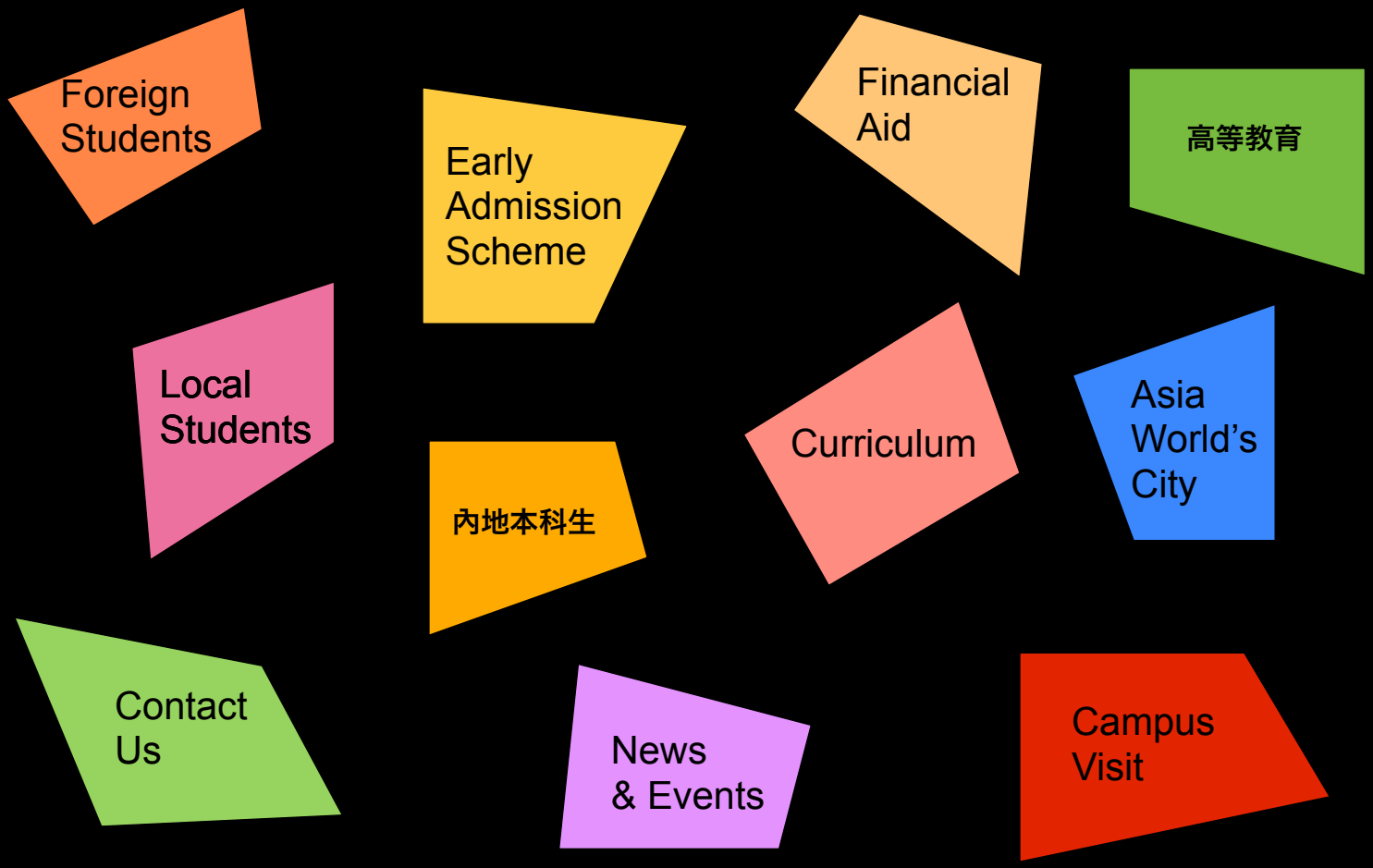


STRUCTURE PLANE

Source: Elements of User Experience
by Jesse James Garrett

Information architecture combines the art and science of structuring, organizing, labeling, navigating, and searching information space.

How do you organize this?



Content Organization Scheme

Logical Grouping

- **Alphabet**
- **Time**
- **Place**
- **Hierarchy**
- **Category**

Source: Richard Wurman, “Information Anxiety”

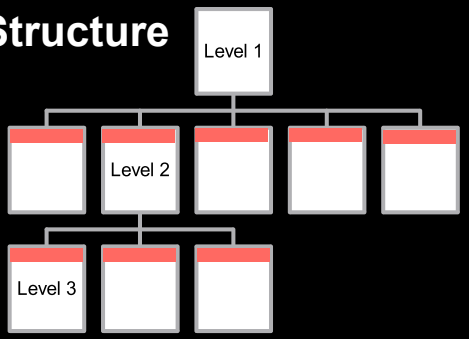
Visual Grouping

- **Linear**
- **Parallel**
- **Web**
- **Matrix**
- **Hierarchical**
- **Spatial Zoom**
- **Overlay**

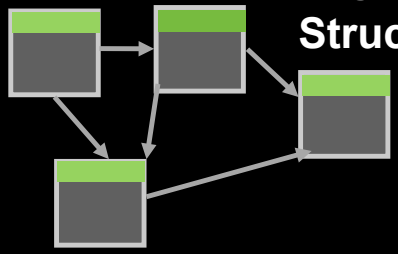
Source: Clement Mok, “Designing Business”

Information Architecture

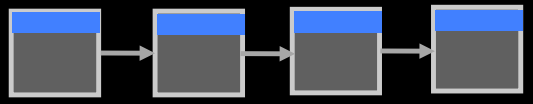
Hierarchical Structure



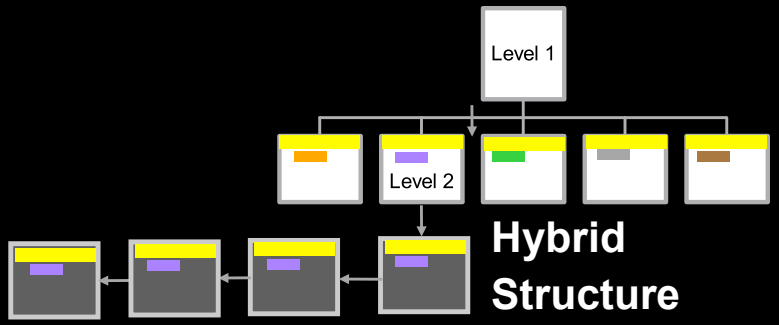
Organic Structure



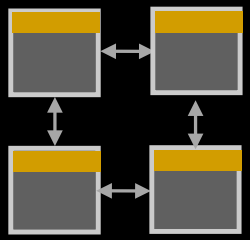
Sequential Structure



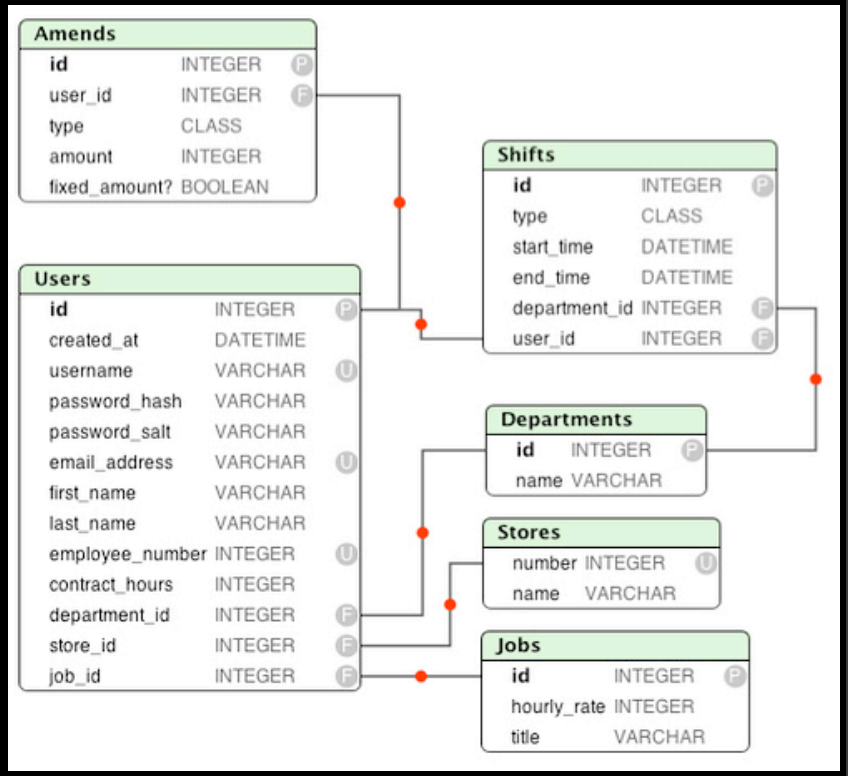
Hybrid Structure



Matrix Structure



Database can be defined as a collection of tables and table can be defined as collection of records. Each record is an unique entry of an object or event, which is made up of attributes:

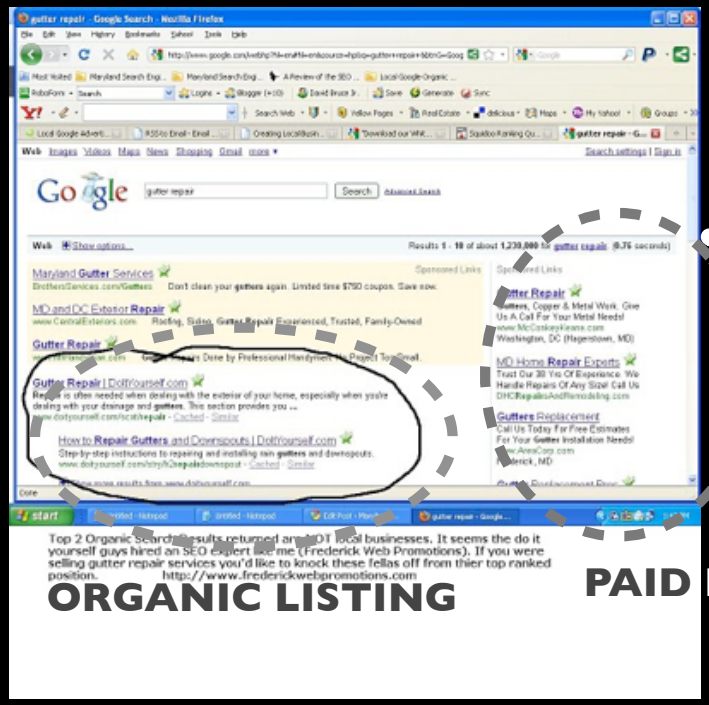


Source:Flickr

- ✓ Most important attribute is a key identifier.
- ✓ There are primary key and secondary key.
- ✓ Information between tables are related by keys.
- ✓ Data about data is called meta-data (e.g. Thesaurus, glossary, controlled vocabulary which can cross-referenced the database)

4

SEARCH INTERFACE



5

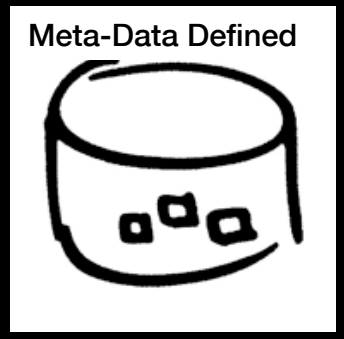
QUERY ENGINE SEARCH INDEX

CRAWLER/SPIDER/ WEB BOT

1

2

CRAW THE WEB



3

INDEXER BUILD UP WEB PAGE INDEX DATABASE

6

RESULT LISTINGS

Question: Can you give me an example of what cannot be easily searched by web spider/crawler?

What is Web Analytics?

Web-based Tool Used for Analyzing Traffic to Your Site:

- What pages have been visited? How many times?
- Where your site visitors come from? Who are they?
- When do they visit? How frequent?
- How long do they stay? What do they buy?
- Where do they stop ? How frequent ?

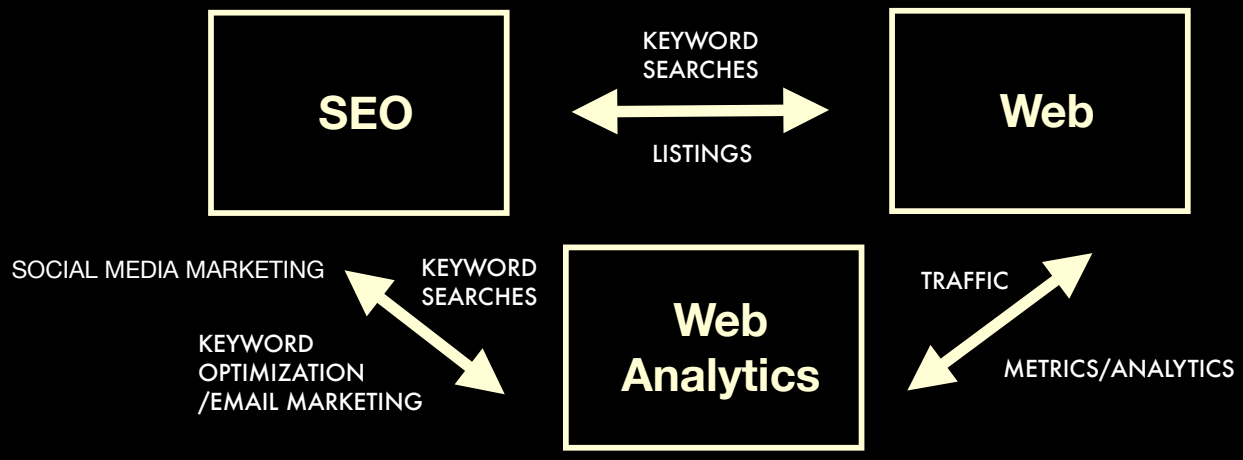
Organic and Paid Listings

- Organic (Natural) Listing (Content Driven - Page Rank Algorithm)
  - Tags (Meta, Title)
  - Body content
  - Links to you (negotiate partnership, build numerous websites, blogs linking to yourself)
- Paid Listing (Pay Per Click:view-->click-->transaction)
  - Sponsored ads
  - Conversion rates (browser->member->customer)

Why Web Analytics?

- Feedback to improve web design
- Know whether the content is enticing
- Has the user experience been captured?
- Who are the real users?
- What do they really care?
- How to make them come back for more?
  - conversion ratio (visitor-->member-->customer)

SEO and Web Analytics



- Web analytics tell you whether your web pages attract traffic (eye balls) or not.
- With feedbacks provided by web analytics, you make changes to your web pages, site structure, title, meta-tag, key word, and link strategies.

Keyword	Search Popularity	Relevance	Competition	Landing Page/ Section

Source: Search Engine Optimization An Hour a Day by Jennifer Grappone and Gradiva Couzin

WHY DO WE NEED CONTENT MANAGEMENT SYSTEM?

Data entities and taxonomy help build the information architecture of the website while the menu items, buttons, and links set the structure for navigation and interaction.

SEO relies on the architecture to build the right content (with proper keywords found in the content Title, Body, Links, and URL alias) to increase web traffic and user engagement!

Web analytics provides the feedback to track the results to see if the content and design is working or not.

END OF MORNING SESSION