

3rd
NJSD

Global Software Development Conference . Nanjing

全球软件大会

2017



人工智能时代的数据工程

MBH CHUANGKE FENG CAO

主讲人：MBH创客蜂巢 CEO 刘端阳

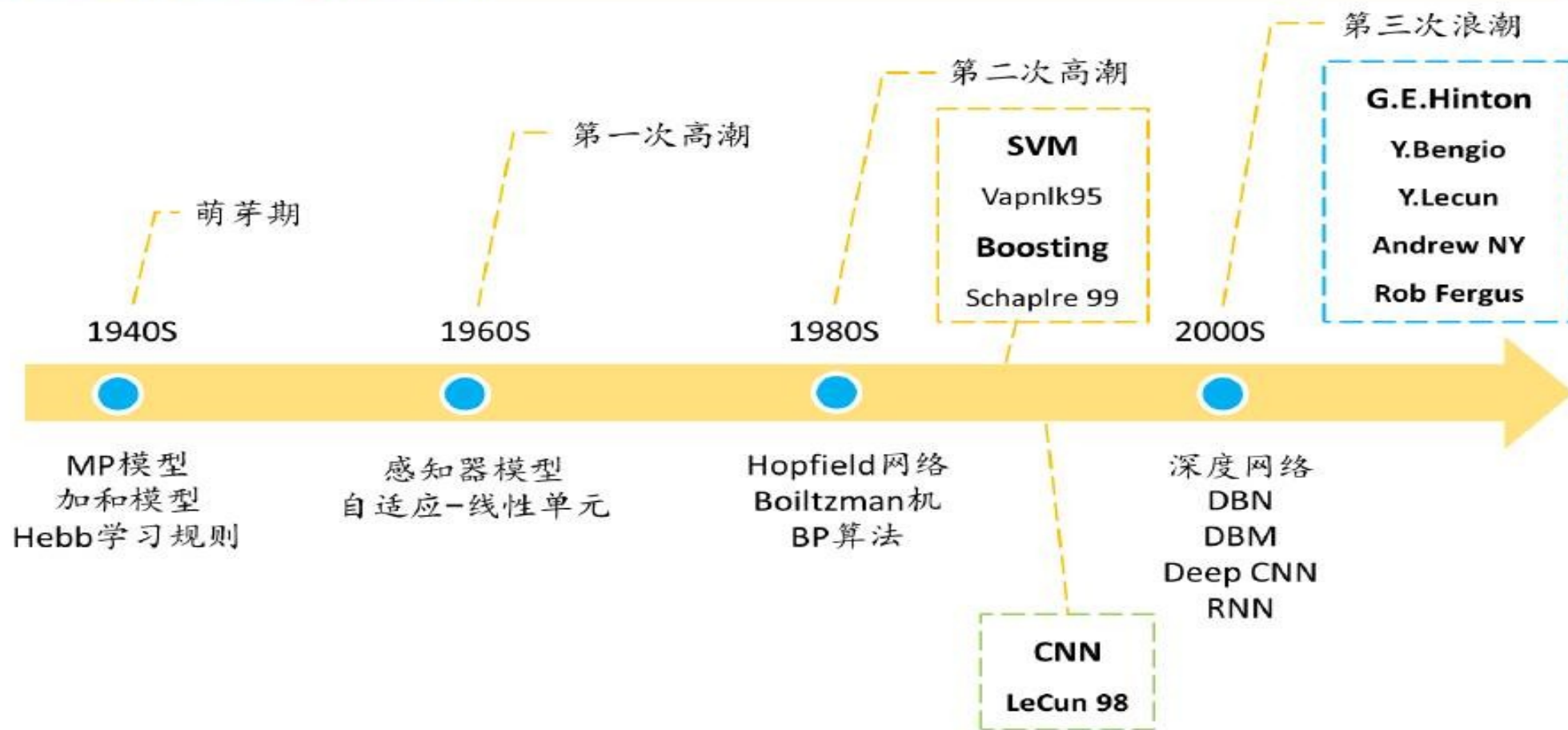
演示提纲

PRESENTATION AGENDA

- 1.AI今年为什么这么火？（数据）
- 2.开放数据运动
- 3.数据驱动的创新
- 4.数据中间件
- 5.众包：群体智慧的典范

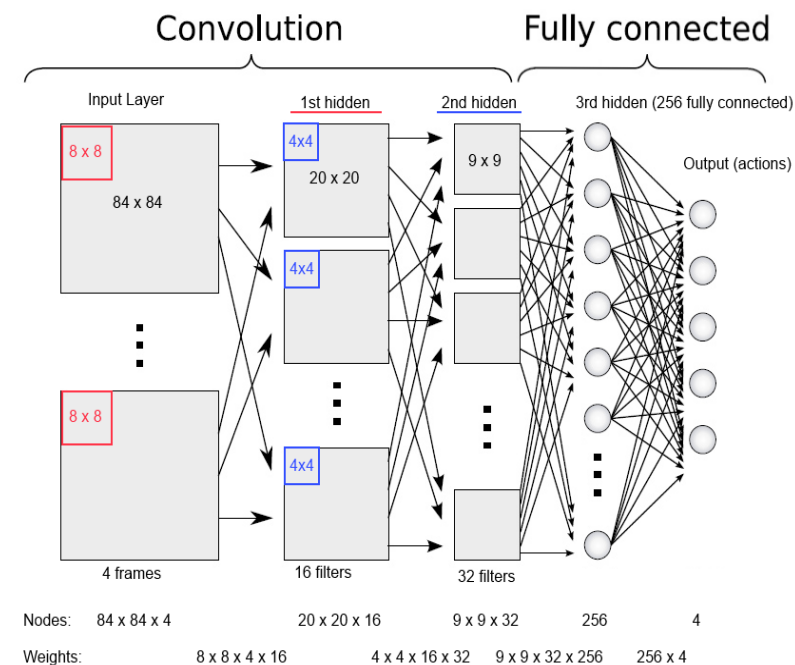
AI为什么火?

图 2: 人工智能发展历程第三次浪潮



资料来源: 招商证券整理

AI为什么火?



Big Data

Cloud Computing

Deep Learning

PC Web

Mobile Web

IOT Data

Perception

Control

Key of AI



◆ Data – Big Data



◆ Performance – Cloud Computing



◆ Model – Deep Learning

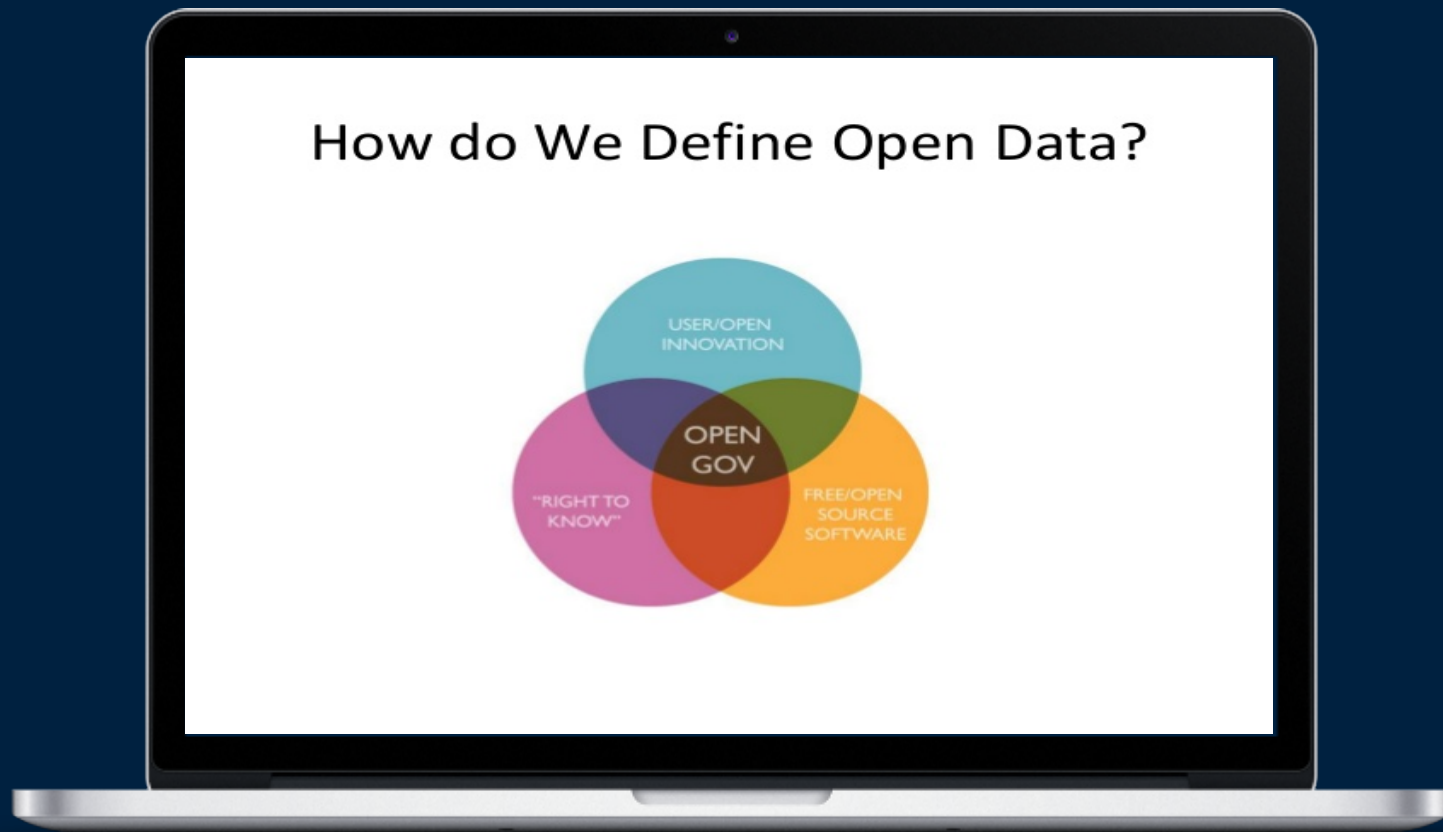


OK, DATA, DATA , DATA



2.1 Define Open Data

Open Data VS Open Knowledge



- ◆ 可获取性
- ◆ 可访问性
- ◆ 再利用和再分发
- ◆ 普遍参与性
- ◆ 拥有二次开发权限

Open Raw Data

Data is the Key , but ...



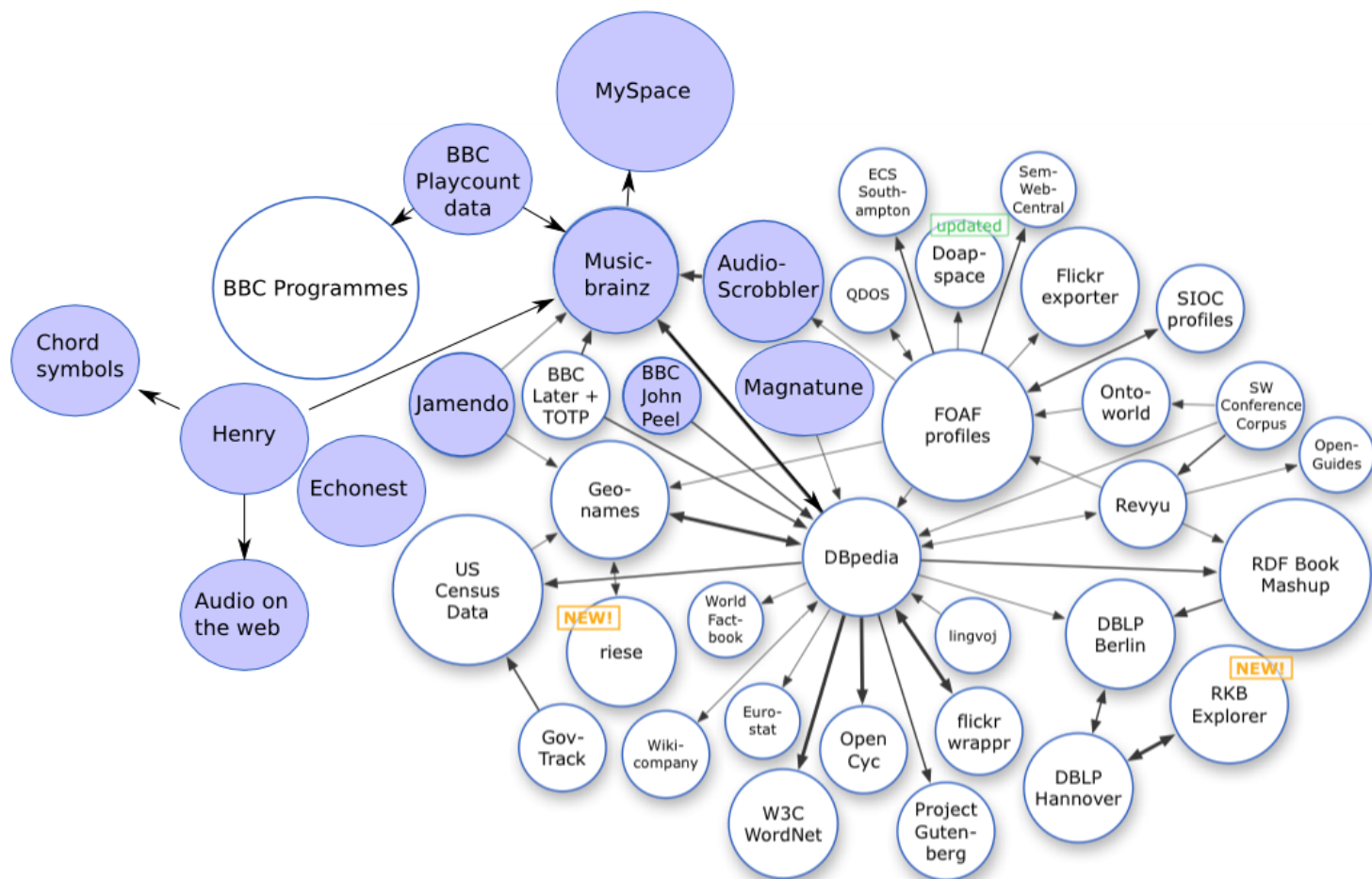
数据私有化



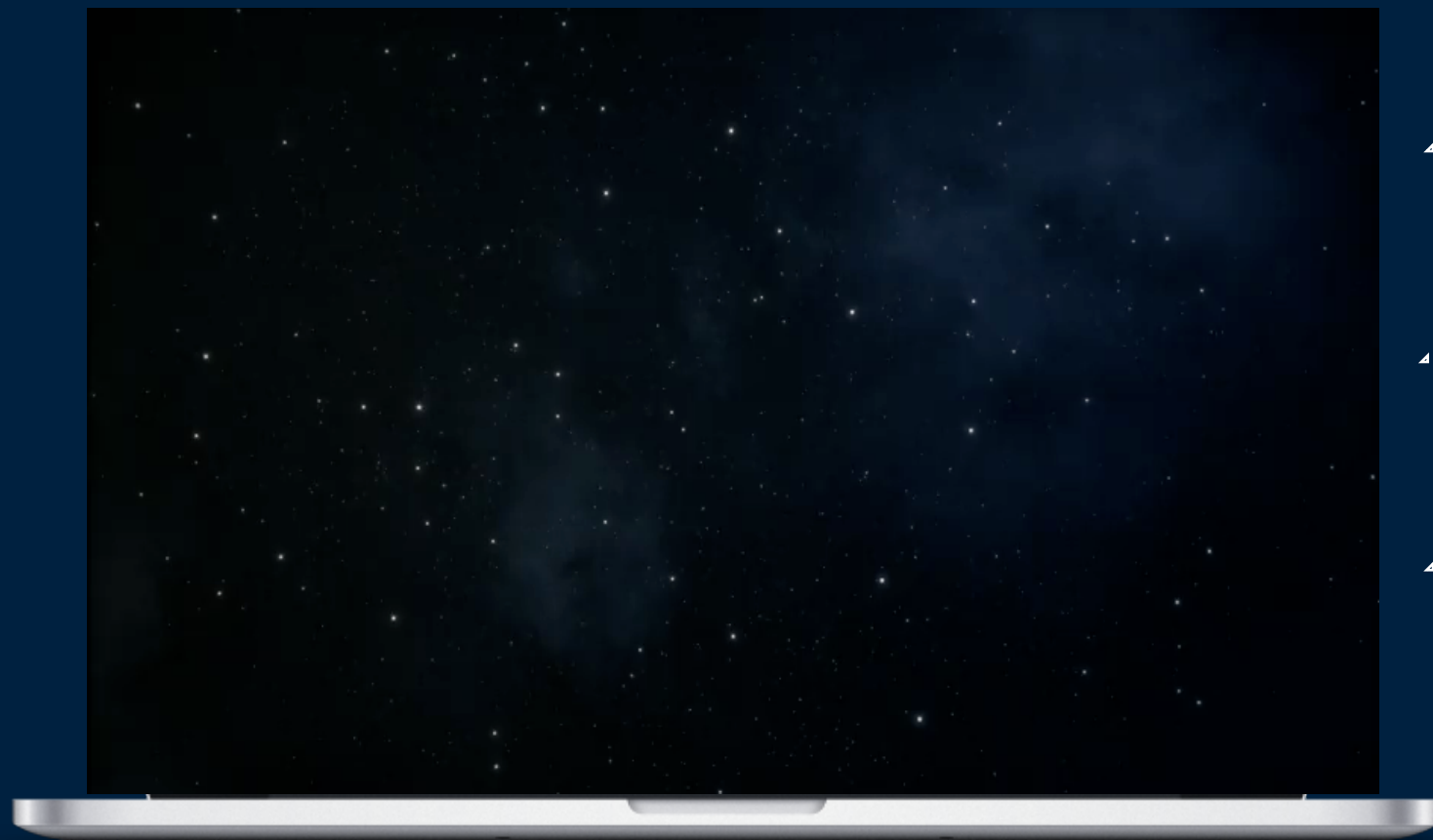
Raw Data被封存



数据被滥用



—• 开放数据驱动创新



Data类型 (gov,science,collaborate...)



Data的属性

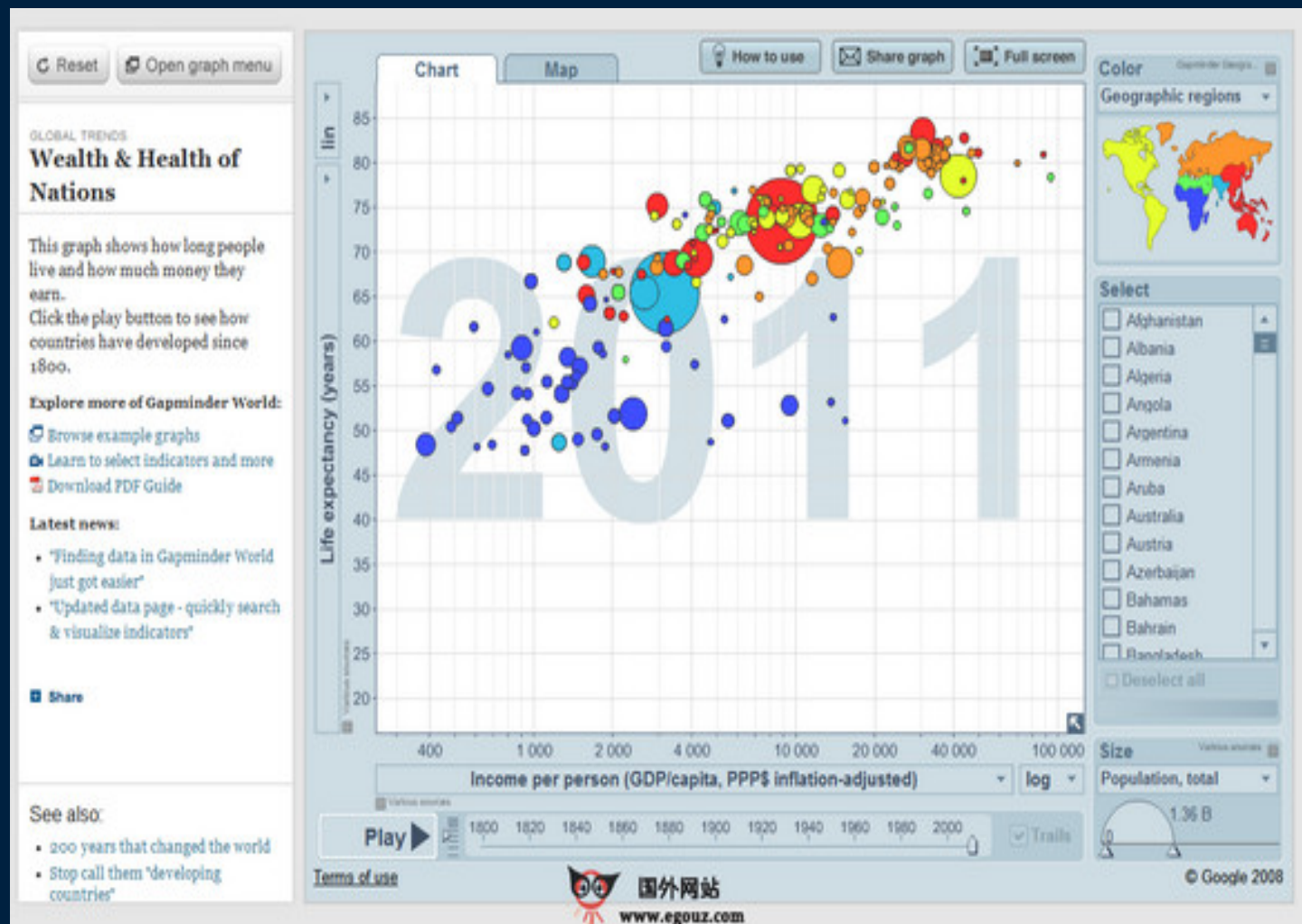


Linked Data: 关联数据将产生巨大的创新

—— 开放数据驱动创新

Open What ?

- 政府共有数据
- 知识data，科学data
- 不对称“信息”数据
- 协作数据：众包之价值



— 开放数据和众包

WHY ?



已有的数据

- a) Document data
- b) Social data

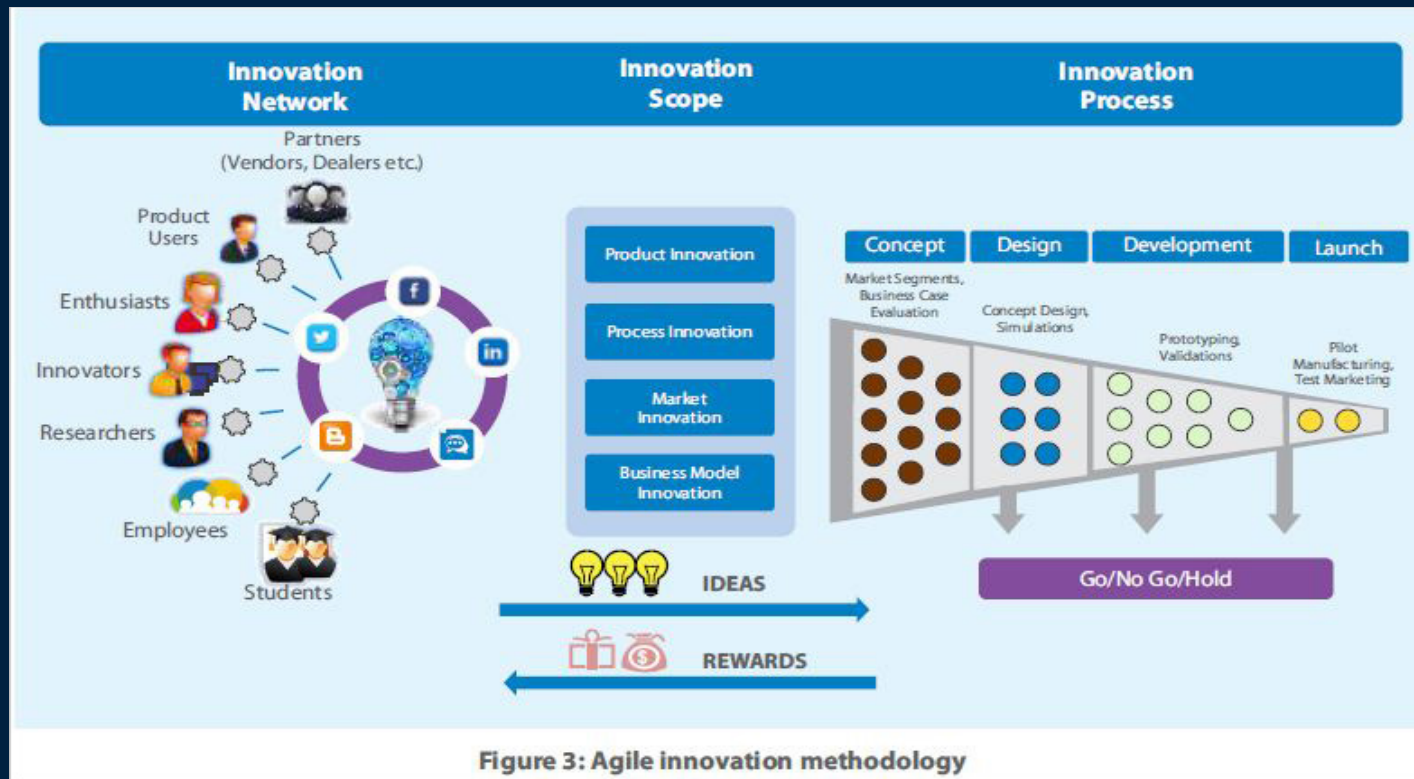


缺失的数据

- a) 协作数据
- b) 任务流本身的数据



众包是open data最好的途径



机器学习：众包信用解决之道



- NP-Hard：亚马逊土耳其机器人



● 众包最大的问题

- 信用
- 不同频的技能:分层
- 接包的质量如何保证: Protocol



机器学习在众包中应用（MBH经验）

- 任务自动分类
- 自动推荐

数据中间件



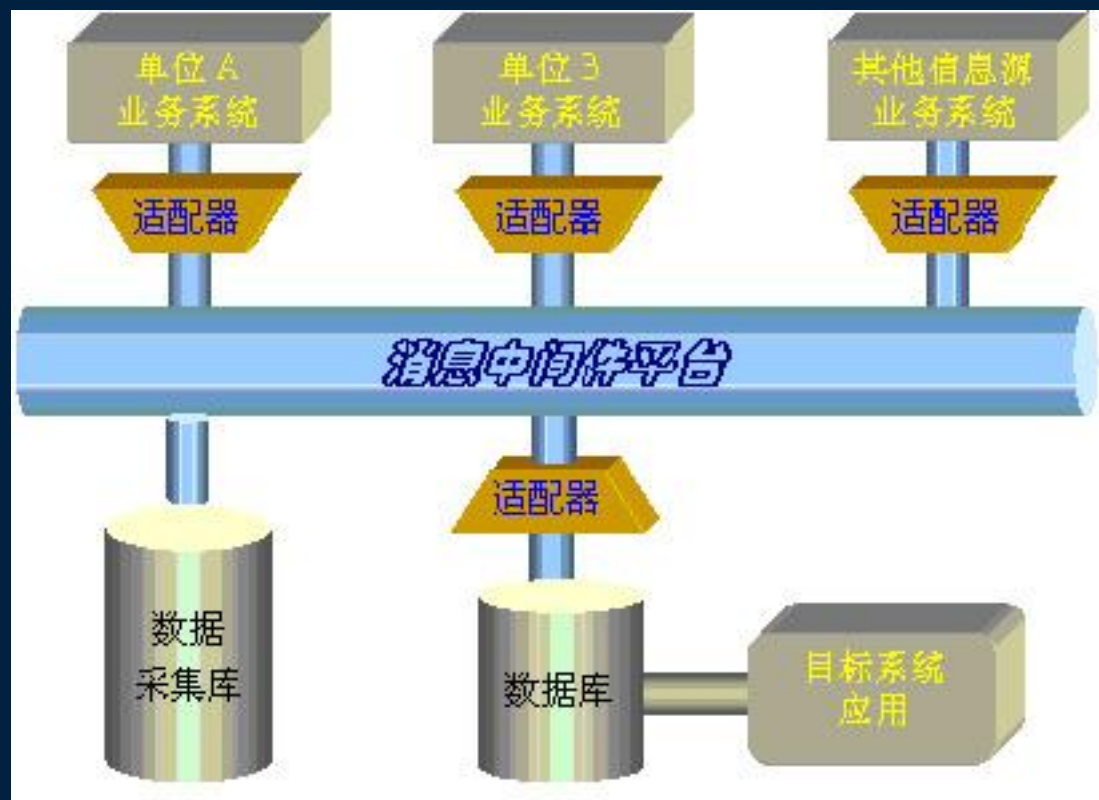
数据中间件是解决开放与隐私的钥匙



Web Services



API for All



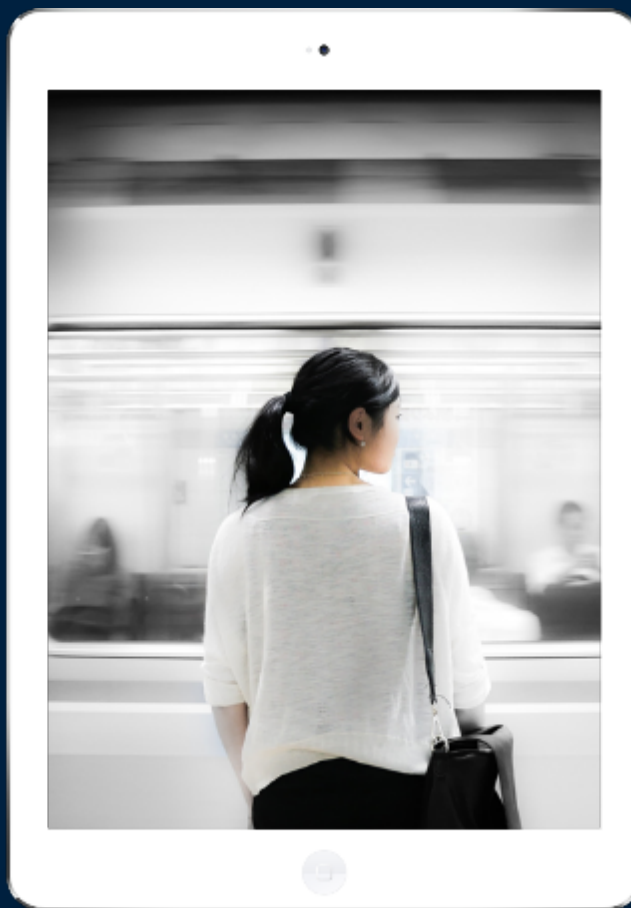
— 数据中间件



中间件的核心



学件
(机器学习专家周志华)



知识中间件 (中科院陆汝)



数据中间件需要哪些技能?

- 【1】Formate of data
- 【2】Python programming
- 【3】RDF三元组进行数据解析

—• 数据中间件



Raw Data



解密



加密



协议



数据损失



基于API或者共同的数据格式 (not RDF...)

谢谢聆听
THANKS FOR YOUR LISTENING!