



中国国际大数据大会
China International Big Data Summit

资料汇编

(仅供内部交流，请勿外传)



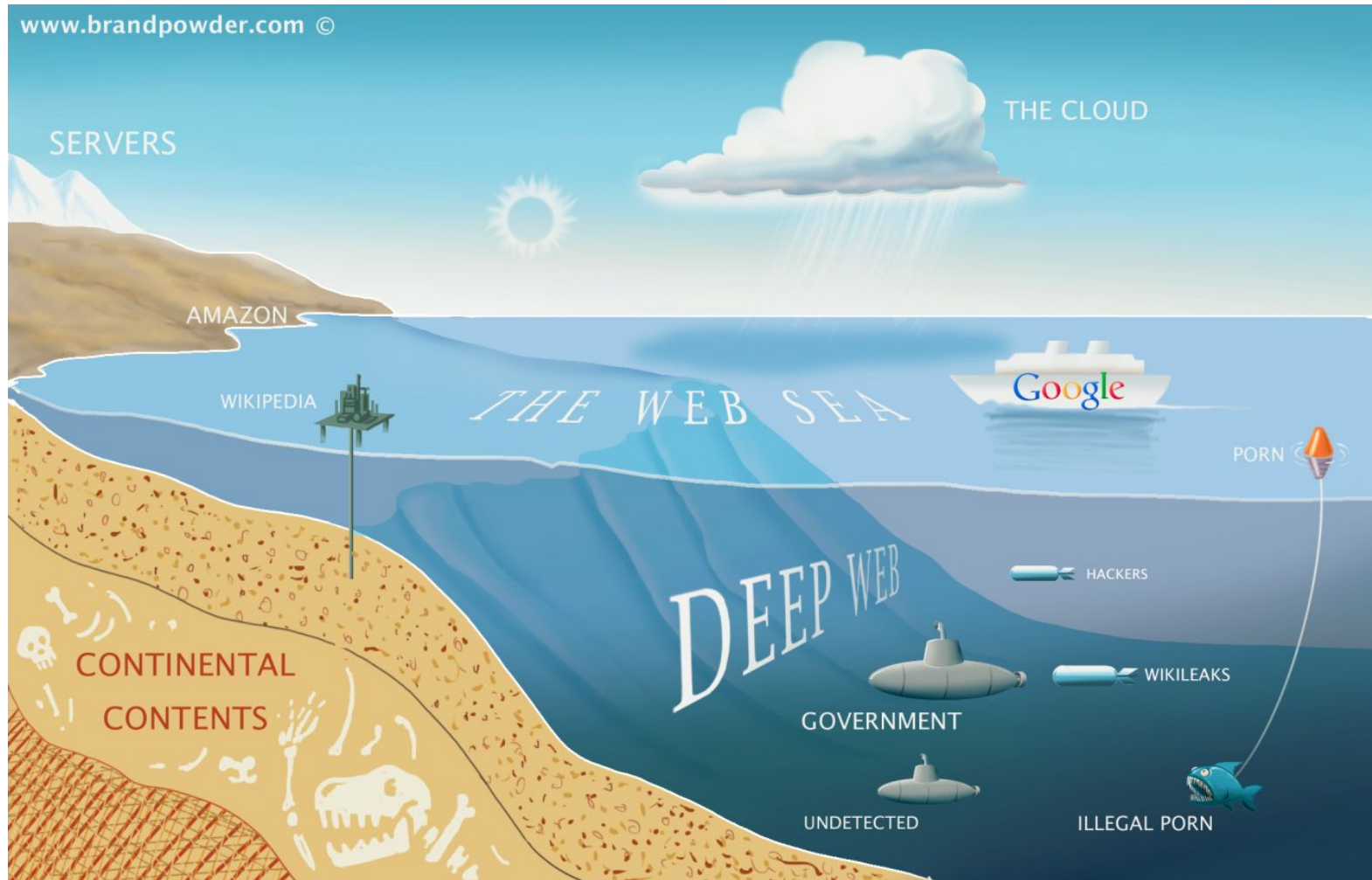
英特尔中国研究院

大数据的开放式创新

@吴甘沙



大数据的暗黑之海



数据之于数据社会，如同
水之于城市，
血液之于身体

数据间的化学作用：大数据 × 行业



乘法效应 + 外部效应

狭义的数据开放

主体：非涉密政府数据、国家拥有的科研数据，部分企业数据

数据开放 不等于 信息公开：原始数据，主动、免费开放

开放标准——5星，数据质量



与其他数据链接

By: Tim Berners-Lee

数据开放的形态



Datahub

易管理性
结构化数据服务
访问和共享的控制
原地可视化和分析



Intel Science and Technology Center for Big Data



数据开放中的问题

数据权属的明晰和敏感数据的界定

数据的脱敏

- 去标识化
- 防止重新标识化：多数据源，统计特性
- 匿名化：K-匿名性，L-多样性，差分隐私

广义数据开放：点对点/多边共享和交易

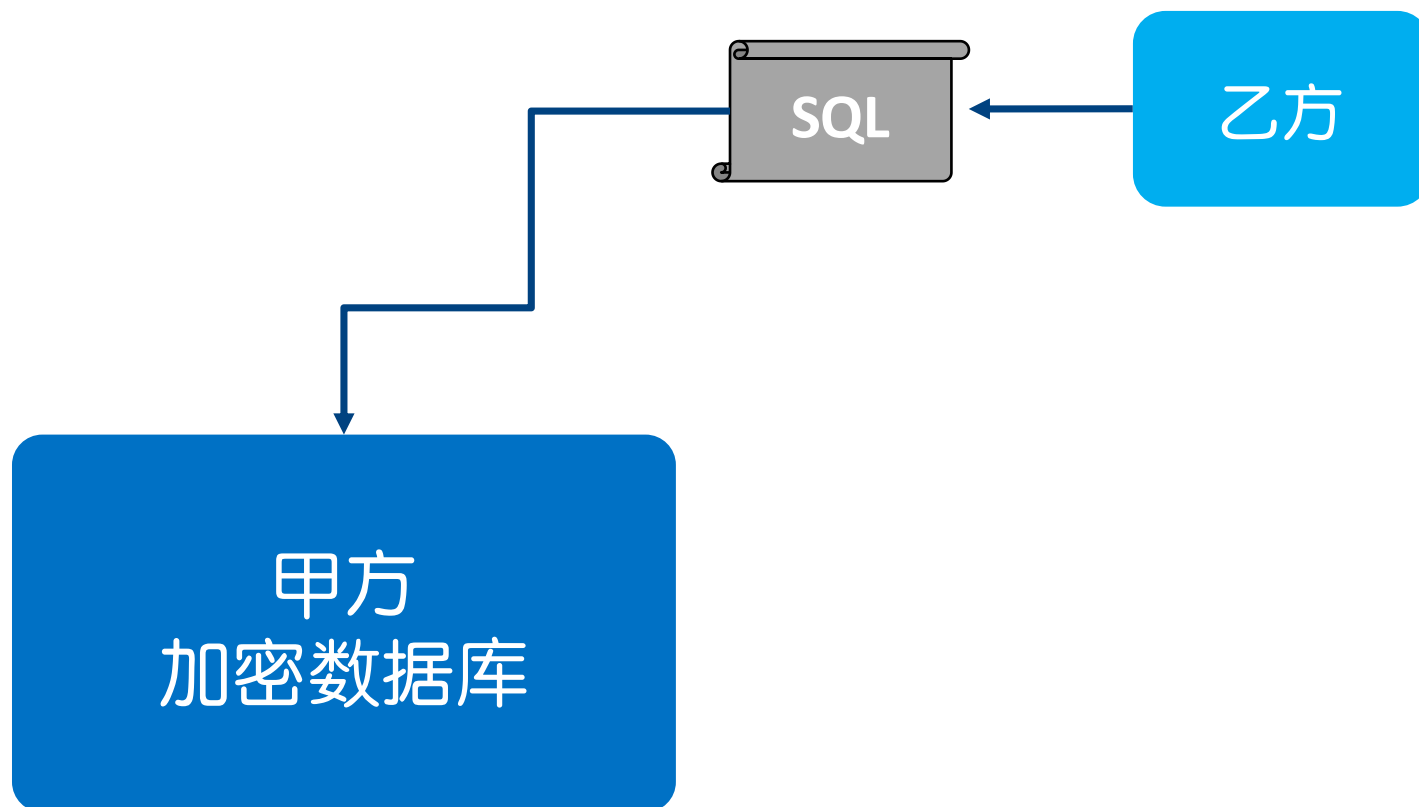
生产资料租赁制 vs. 生产资料所有制

数据权利的保障

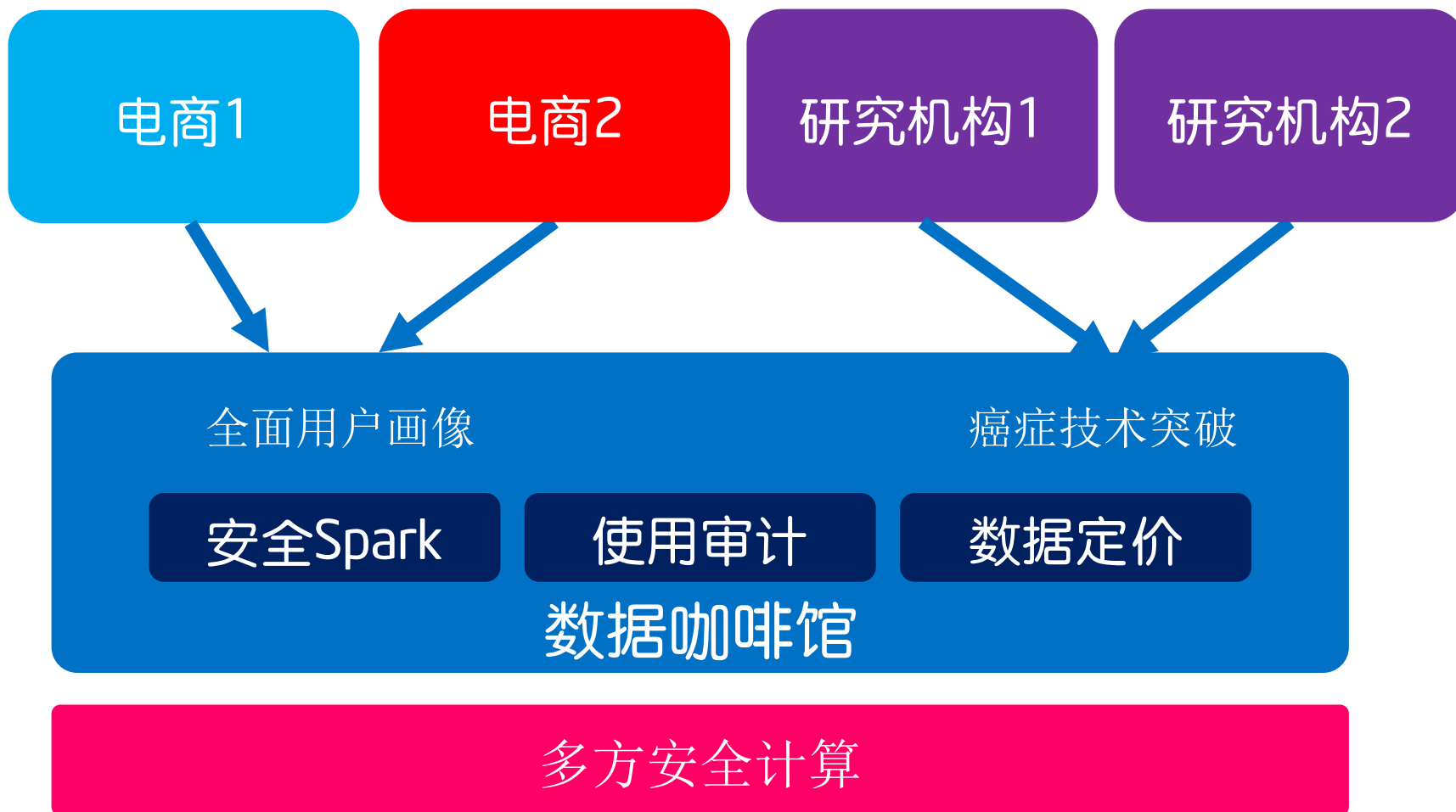
- 可用但不可见
- 使用审计：符合隐私规定，防泄漏
- 数据的定价

从点对点走向多边：数据服务，数据市场和数据交易所

可用而不可见案例1：同态加密数据库



可用而不可见案例2：多方安全计算



数据社会的基础是数据经济

大数据时代的“梅特卡夫定律”是什么？

数据定价和估值

公开市场交易和“数据即服务”中的数据定价

点对点共享中的数据定价

企业数据资产的估值

个人数据的定价

在共享和交易中防止劣质数据

大数据基础设施的开放

云计算是答案

- 平台即服务：大数据基础处理和分析平台

降低数据思维者的门槛，释放他们的创造力

- decide.com
- Prismatic

数据价值提取能力的开放

Analytics as a Service



如Tesco和Dunnhumby



如Palantir



如Kaggle：针对长尾企业

英特尔愿与政府、学术和产业界
同行者同创数据开放、共享和交
易的美好未来

敬谢聆听