

# Dados abertos: uma perspectiva a partir dos Estudos Sociais da Ciência

Maria Cristina Soares Guimaraes

**LICTS/ICICT/FIOCRUZ**

**Rio de Janeiro, setembro de 2017**

# Passos...

---

- O ponto de partida
  - Infraestrutura de dados
  - Big science x Little science
  - Open data para quem?
  - Políticas de incentivo
  - Desafios
-

# Dados abertos?

---

- Perspectiva Estudos de Informação/Estudos Sociais da Ciência;
  - Abertura é um movimento contra crescente comodificação e privatização dos resultados das pesquisas;
  - Acelerar fluxos para acelerar produção de conhecimento (*Knowledge-driven science vs. Theory-driven knowledge*);
  - TICs operam nos dois extremos: oferta e demanda;
  - Sempre um sistema sócio-técnico.
-

# Não perder de vista...

---

- Dados não preexistem a sua geração; eles não surgem do nada, eles não são inevitáveis: protocolos, processos organizacionais, escalas de medição, categorias e padrões são projetados, negociados e debatidos;
- Dados não são pré-analíticos e pré-factuais, ou contrário, são constitutivos das ideias, técnicas, tecnologias, pessoas, sistemas e contextos que os concebem, produzem, processam, gerenciam e analisam;
- Dados são situados, contingentes e relacionais, e são utilizados contextualmente para tentar atingir determinados objetivos e metas.

---

(Bowker and Star 1999; Lauriault 2012)

# Dimensões e contextos

---

## Ético

Quais dados? Como foram gerados? Como são compartilhados, transacionados, protegidos, utilizados? Como os indivíduos são “classificados”?

## Técnico

Como os métodos de captura e medição geram dados corretos, limpos e precisos.

## FRAMING DATA

## Filosófico

A noção de que um dispositivo produz uma visão de mundo objetiva e realista.

## Político e Econômico

Visão como commodity (normatização, regulação, financiamento e negociação).

## Tempo e Espaço

Dados e seus arranjos possuem histórias e geografias – como são processados e analisados varia ao longo do tempo e do espaço  
Kitchin, 2014

# Infraestruturas de dados

---

Arranjos de dados e repositórios não são técnicos e instrumentais em sua natureza; antes, são expressões de conhecimento e poder. Seus processos de organização envolvem negociações, tensões e contingências, refletem racionalidades acerca do entorno e do ambiente externo, reforçando-as.

Como para Foucault (1977), são como um dispositivo (discursos, instituições, arquiteturas, decisões regulatórias, proposições filosóficas, declarações científicas etc.).

---

# Aparatos e elementos que orientam as infraestruturas

(KITCHIN & LAURIAULT, 2014)

<b>Apparatus</b>	<b>Elements</b>
Systems of thought	Modes of thinking, philosophies, theories, models, ideologies, rationalities, etc.
Forms of knowledge	Research texts, manuals, magazines, websites, experience, word of mouth, chat forums, etc.
Finance	Business models, investment, venture capital, grants, philanthropy, profit, etc.
Political economy	Policy, tax regimes, incentive instruments, public and political opinion, etc.
Governmentalities and legalities	Data standards, file formats, system requirements, protocols, regulations, laws, licensing, intellectual property regimes, ethical considerations, etc.
Materialities and infrastructures	Paper/pens, computers, digital devices, sensors, scanners, databases, networks, servers, buildings, etc.
Practices	Techniques, ways of doing, learned behaviours, scientific conventions, etc.
Organisations and institutions	Archives, corporations, consultants, manufacturers, retailers, government agencies, universities, conferences, clubs and societies, committees and boards, communities of practice, etc.
Subjectivities and communities	Of data producers, experts, curators, managers, analysts, scientists, politicians, users, citizens, etc.
Places	Labs, offices, field sites, data centres, server farms, business parks, etc. and their agglomerations
Marketplace	For data, its derivatives (e.g., text, tables, graphs, maps), analysts, analytic software, interpretations, etc.

# Dimensões e contextos

---

"Databases create silences, adding to the inherent gaps in the data, as much they reveal relationships between data and enable some questions to be answered; they constrain and facilitate through their ontology, producing various presences and absences of relations" (Ruppert, 2012).

---



# QUANDO RELACIONADOS AO MUNDO ACADÊMICO

---

Sistemas sócio-técnicos : Infraestruturas de  
Conhecimento/Culturas Epistêmicas/Domínios de  
Conhecimento

Redes de instituições, micropolíticas, atores,  
artefatos que produzem e compartilham  
conhecimento

Como evoluem? Como distribuem autoridade,  
influência e poder?

---

# Big Science x Little Science

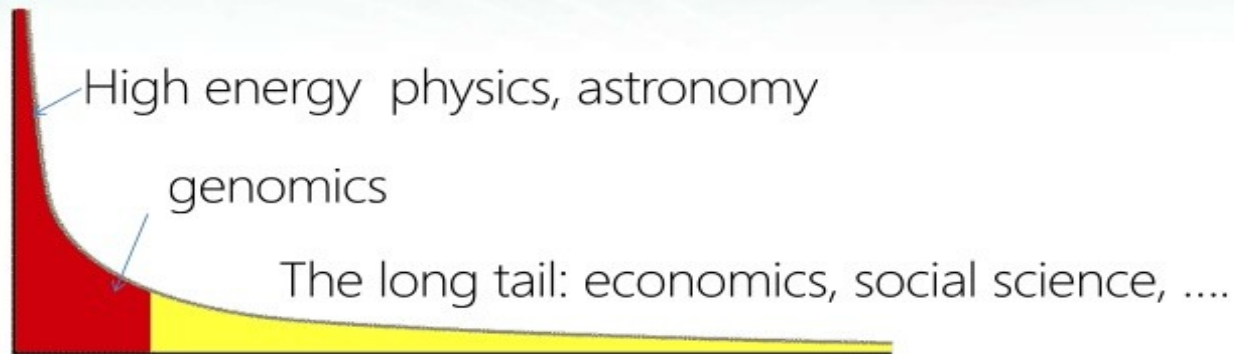
---

Diferenças qualitativas entre diferentes domínios de conhecimento:

- Big science: ampla dimensão e capilaridade da rede de atores; forte apoio das TICs para produção e processamento de dados; rede informal bem estruturada e concordância metodológica;
  - Little Science comunidades mais restritas; menor volume de dados e uso restrito, caráter local, menor variedade metodológica, relativamente pouca tecnologia.
-

# Big Science x Little Science

## The Long Tail of Science



Collectively "long tail" science is generating a lot of data  
Estimated at over 1PB per year and it is growing fast.

80-20 rule: 20% users generate 80% data but not necessarily 80% knowledge

CSTI Meeting. October 2012  
Dennis Gannon



<https://portal.futuregrid.org>

# Acesso dados abertos

---

Diferença acesso literatura X acesso dados

- No geral, políticas de acesso aos dados abertos está atrelada àquelas de acesso à literatura;
  - Desafios diferentes: caráter proprietário muito mais difuso e difícil de determinar, especialmente direitos e responsabilidades (p.e., em colaborações);
  - Editoras publicam periódicos; como publicar dados???  
Liberação conduzida pelo pesquisador;
  - Modelo de governança: como garantir equidade, eficiência e sustentabilidade?
  - Problemas de acesso: formas diferenciadas de representação
-

# If We Share Data, Will Anyone Use Them?

Wallis et al., 2013

---

- Ethos científico (Robert Merton, 1942) – abertura para escrutínio (comunalismo e ceticismo organizado) ;
  - Abertura traz vantagens quali e quanti para dinâmica da ciência (replicação sem duplicação e integridade) – Ética
  - E a prática? Muito pouco praticado pelos pesquisadores *Little science* (ausência infraestrutura apropriada; preocupação com direitos de propriedade, incentivo e investimento pessoal para gerir e preservar os dados de forma a resguardar contexto de produção)
-

# Desincentivos na Little science

---

- Sistema de recompensa – reputação e incentivos desalinhados
  - Acesso e disponibilidade
-

# POLÍTICAS PARA DADOS ABERTOS???

---

- Benefícios para reputação (não somente publicação), investir em curadoria – melhores conjuntos de dados;
- Reduzir custos de transação – aumentar valor da reprodutibilidade (editoras e agencias de financiamento)
- Autoria duvidosa – direitos ficam difusos
- Transparência (facilidade de acesso) – repositórios
- Modelo de governança

---

INCENTIVOS INSTITUCIONAIS, DISCIPLINARES E SISTÊMICOS

# DO SMALL PARA BIG DATA - DESAFIOS CANDENTES

---

*Data science*: crescente interdisciplinaridade e complementaridade entre os campos disciplinares

- Impredicabilidade (impossibilidade de antecipar consequências);
- Natureza coletiva e distribuída do *reasoning* e agência – produzido localmente, integrado e interpretado globalmente;
- *Path-dependency* (“lock-in” na infraestrutura tecnológica e na escolha de algoritmos, terminologias e padrões).



# Seguindo adiante

---

“We need to open a discourse where there is no effective discourse now about the varying temporalities, spatialities and materialities that we might represent in our databases, with a view to designing for maximum flexibility and allowing as possible for an emergent polyphony and polychrony. Raw data is both an oxymoron and a bad idea; to the contrary, data should be cooked with care”.

---

Geoffrey Bowker (2005, p. 183-184)