

Dados de investigação e arquivos em ciências sociais:

contributos do projeto CESSDA SaW

Pedro Moura Ferreira
ICS-UL

CESSDA – Consortium of European Social Sciences Data Archives

NATIONAL DATA SERVICES

Find your National Data Service



Objetivos do projeto SaW: Strengthening and Windening

- Estado da arte dos serviços proporcionados pelos arquivos
- Necessidades atuais e futuras dos investigadores
- Quatro domínios de dados explorados:
 - Académico
 - Saúde
 - Estatísticas oficiais
 - História

Domínio acadêmico - Situação

- Atividade central dos arquivos da rede CESSDA
- Tipos diferentes de dados em termos de preservação, curadoria e gestão.
- Dados de inquéritos tem sido o alvo preferencial. A emergência de novos tipos de dados (ex: big data, administrativos, etc.) e produtores (INE, Governo, Bancos) desafiam as capacidades tradicionais dos arquivos.
- Além de dados quantitativos, surgem os qualitativos: transcrições de entrevistas, imagens ou gravações.
- A investigação qualitativa menos habituada à análise secundária e à reutilização de dados produzidos por outros
- Questões de preservação, legais e éticas colocam importantes desafios à ao tratamento dos dados qualitativos.

Domínio académico - Recomendações

- Adequar a infraestrutura para lidar com os dados qualitativos
- Promover uma política de coleções de dados qualitativos:
 - Identificar a procura para dados qualitativos
 - Desenvolver instrumentos e serviços especializados
 - Sensibilizar para a sua entrega nos arquivos
 - Estabelecer a cooperação com os atores da investigação qualitativa

Domínio da saúde - Situação

- Suscitam o interesse dos arquivos porque a investigação é essencial para as políticas de saúde
- Questões legais permanecem: a anonimização pode levar a perda de informação relevante
- Necessidade de metadados que minimizem as diferenças entre diferentes standards.

Domínio da saúde: Recomendações

- Desenvolver a infraestrutura para permitir a receção, curadoria e disseminação dos dados de saúde
- Estabelecer acordos com produtores de dados de saúde (institutos / organizações)
- Colaborar com outras instâncias no sentido de desenvolver uma moldura legal que proteja a informação pessoal e permita, se e quando possível, a investigação

Domínio das estatísticas oficiais - Situação

- As estatísticas produzidas por variados atores (governamentais, reguladores, privados)
- Web data ou dados transacionais, além de novos dados administrativos, são outras fontes a considerar já que suscitam interesse da investigação
- Big data pode proporcionar outros recursos em larga escala
- Há cooperação entre arquivos e as estatísticas oficiais.
- Limitações: no portal da pesquisa, no acesso aos dados administrativos, no uso dos metadados
- Os investigadores apontam ainda limitações no acesso aos dados dos Institutos Nacionais de Estatística (ex.: tempo de demora; extensão e compatibilidade internacional dos metadados)

Domínio das estatísticas oficiais - Recomendações

- Desenvolver e promover big data, articulando a análise e as técnicas das ciências sociais com as ciência computacional, pelas implicações nas orientações políticas das sociedades participativas e inclusivas.
- Estabelecer colaborações com outras instituições em ordem a desenvolver metodologias e instrumentos que facilitem e promovam a investigação.

Domínio da História: - Situação

- Interesse periférico na medida em que se insere mais nas humanidades do que nas ciências sociais. Há arquivos que proporcionam acesso e disseminam dados históricos.
- Há novas tendências na investigação: história oral, séries temporais. Maior proximidade entre as bases de dados sociais e históricos
- As descrições de metadados podem ser insuficientes na medida em que os dados não foram recolhidos com propósito estatístico.

Domínio da História: - Recomendações

- Estabelecer cooperação com outros atores (investigação e arquivos)
- Priorizar a recolha, curadoria e disseminação de séries temporais
- Desenvolver descrições de metadados e seguir os desenvolvimentos tecnológicos em ordem a incentivar outras abordagens às bases relativas a dados históricos.

Conclusões

- As fontes de dados são imensas e crescentes
- A resposta dos arquivos é tardia e insuficiente
- Os recursos humanos e tecnológicos necessitam de desenvolvimentos
- As necessidades da investigação são fundamentais para criar pressão nas respostas a implementar