

Transparência dos (ex-)microdados do Inep

Fernanda Campagnucci | diretora executiva



Audiência Pública
Comissão de Educação e Cultura
Senado Federal
26.10.2023

Questões

que abordaremos nesta apresentação

1. TOMADA DE DECISÃO

O que aconteceu à época:
considerações sobre a forma e o
mérito da decisão de agregar os
microdados da educação

2. IMPACTO DA DECISÃO

O que aconteceu de lá pra cá: o
que perdemos com a agregação
dos dados, comentários sobre a
estratégia da sala segura e sobre o
acesso aos dados atualmente

3. PERGUNTAS SEM RESPOSTAS (PÚBLICAS)

Algumas questões que deixamos
para que o debate seja
aprofundado

microdados

definições do [Cetic.br](#)
e do [Banco Mundial](#)

“A menor fração de um dado coletado em uma pesquisa. Retrata [...] a resposta individual dos informantes ao questionário aplicado”

“Dados no nível da unidade de análise obtidos em questionários, censos e sistemas administrativos. [...] permitem compreensão aprofundada de questões socioeconômicas e das interações entre fenômenos.”

Os censos publicados pelo
Inep não estão mais
detalhados como microdados

Censo Escolar da Educação Básica
Censo da Educação Superior

Tomada de decisão

Questões sobre a forma

- **Falta de diálogo:** a publicação de microdados era um processo consolidado há mais de uma década e com amplo ecossistema de usuários; precisava ter havido método de escuta para articular riscos, benefícios e medidas de mitigação
- **Opacidade:** estudo e pareceres pela mudança foram divulgados apenas após repercussão social negativa



Tomada de decisão

Questões sobre o mérito

- **Fragilidades:** por importante que seja, estudo que embasou decisão não poderia ser o único - abrange apenas uma parcela da questão, a segurança.
- **Comparação prejudicada:** Compara Brasil com três países e usa referências anedóticas. Usa exemplo do Censo dos EUA, mas lá a mudança foi anunciada em 2017 e passou por amplo debate público até 2020; na Austrália, dados são fornecidos mediante pagamento, etc.
- **Métodos de anonimização:** foram tecnicamente os melhores? Atendem ao interesse público sobre esses dados?

Impactos do fechamento

Menos dados, menos possibilidades de análise

Censo da Educação Superior

ANTES

| | | | | | |
|---|---------------------------------|------------|-----|--------------|-----|
|  DM_ALUNO_2019 | Planilha OpenOffice.org XML 1.0 | 530.728 KB | Não | 3.022.144 KB | 83% |
|  DM_CURSO_2019 | Planilha OpenOffice.org XML 1.0 | 1.608 KB | Não | 9.611 KB | 84% |
|  DM_DOCENTE_2019 | Planilha OpenOffice.org XML 1.0 | 13.948 KB | Não | 48.371 KB | 72% |
|  DMIES_2019 | Planilha OpenOffice.org XML 1.0 | 224 KB | Não | 696 KB | 68% |

- Uma planilha com **dados por aluno**, permitindo cruzamentos por **raça-cor, sexo, curso e local**, por ex.
- Uma planilha com informações sobre **cada um dos cursos regulares** oferecidos no Brasil
- Uma planilha com informações sobre os **docentes**
- Uma planilha com informações sobre as **universidades**

Censo da Educação Superior

DEPOIS

|  | CADASTRO_CURSOS_2019 | Planilha OpenOffice.org XML 1.0 | 11.803 KB | Não | 130.795 KB | 91% |
|---|---------------------------------|---------------------------------|-----------|-----|------------|-----|
|  | md5_microdados_ed_superior_2019 | Planilha OpenOffice.org XML 1.0 | 1 KB | Não | 1 KB | 18% |
|  | MICRODADOS_ED_SUPIES_2019 | Planilha OpenOffice.org XML 1.0 | 301 KB | Não | 970 KB | 70% |

- Uma planilha com os **dados dos cursos** (a antiga DM_CURSO_2019), com diversas colunas extras que **resumem** a antiga tabela DM_ALUNO_2019 informando **quantidades de alunos por curso** - totais, por forma de seleção e por tipo dentro de uma categoria (uma coluna dizendo quantos alunos brancos, outra quantos alunos negros, outra quantos alunos homens, outra quantas alunas mulheres...)
- Uma planilha com informações sobre as universidades (antiga DMIES_2019)
- **Não há mais informação sobre os docentes**, além da mera quantidade

Educação Básica

ANTES

- Tabelas sobre docentes, por região, incluindo dados sobre a adequação entre sua **formação e a disciplina** que lecionam
- Tabela com dados básicos sobre as **escolas**
- Tabela sobre os **gestores escolares**
- Tabelas com dados de **alunos matriculados**, por região
- Tabela com **dados de turmas** - cada aluno e professor era vinculado a uma turma e cada turma a uma escola

| | | |
|---|---------------------------------|------------|
|  DOCENTES_CO | Planilha OpenOffice.org XML 1.0 | 18.196 KB |
|  DOCENTES_NORDESTE | Planilha OpenOffice.org XML 1.0 | 67.021 KB |
|  DOCENTES_NORTE | Planilha OpenOffice.org XML 1.0 | 21.017 KB |
|  DOCENTES_SUDESTE | Planilha OpenOffice.org XML 1.0 | 92.145 KB |
|  DOCENTES_SUL | Planilha OpenOffice.org XML 1.0 | 38.333 KB |
|  ESCOLAS | Planilha OpenOffice.org XML 1.0 | 12.965 KB |
|  GESTOR | Planilha OpenOffice.org XML 1.0 | 7.176 KB |
|  MATRICULA_CO | Planilha OpenOffice.org XML 1.0 | 182.735 KB |
|  MATRICULA_NORDESTE | Planilha OpenOffice.org XML 1.0 | 751.735 KB |
|  MATRICULA_NORTE | Planilha OpenOffice.org XML 1.0 | 242.159 KB |
|  MATRICULA_SUDESTE | Planilha OpenOffice.org XML 1.0 | 926.671 KB |
|  MATRICULA_SUL | Planilha OpenOffice.org XML 1.0 | 332.468 KB |
|  MD5_microdados_ed_basica_2019 | Planilha OpenOffice.org XML 1.0 | 1 KB |
|  TURMAS | Planilha OpenOffice.org XML 1.0 | 43.448 KB |

Educação Básica

- Restou apenas uma **tabela de escolas**, com colunas informando **quantas turmas e quantos alunos existem por nível**, com quebras demográficas.
- Perdeu-se todo o **detalhamento sobre docentes**, não existindo sequer os quantitativos por gênero ou disciplina

DEPOIS

| | | |
|---|---------------------------------|-----------|
|  md5_micromdados_ed_basica_2019 | Planilha OpenOffice.org XML 1.0 | 1 KB |
|  micromdados_ed_basica_2019 | Planilha OpenOffice.org XML 1.0 | 24.819 KB |

ENADE

Outro tipo de caso: a perda de informação pela fragmentação excessiva

ANTES

| | | | | | |
|---|---------------------------------|-----------|-----|------------|---|
|  microdados_enade_2018 | Planilha OpenOffice.org XML 1.0 | 44.274 KB | Não | 254.526 KB | 8 |
|---|---------------------------------|-----------|-----|------------|---|

- Uma tabela que informava os dados do resultado da prova por aluno que o respondeu

ENADE

- **32 tabelas**, cada uma com **parte das variáveis**
- Não é garantido que as informações constantes numa linha de uma tabela, como a **nota geral**, se refiram ao mesmo estudante que a informação da mesma linha em **outra tabela**, como a **declaração de cor e raça**.
- Com isso, **ficam prejudicadas análises da desigualdade na saída do ensino superior**

DEPOIS

| | | | | |
|--|---------------------------------|-----------|-----|------------|
|  microdados2018_arq1 | Planilha OpenOffice.org XML 1.0 | 187 KB | Não | 22.696 KB |
|  microdados2018_arq2 | Planilha OpenOffice.org XML 1.0 | 2.083 KB | Não | 13.073 KB |
|  microdados2018_arq3 | Planilha OpenOffice.org XML 1.0 | 26.213 KB | Não | 141.296 KB |
|  microdados2018_arq4 | Planilha OpenOffice.org XML 1.0 | 8.367 KB | Não | 49.040 KB |
|  microdados2018_arq5 | Planilha OpenOffice.org XML 1.0 | 1.408 KB | Não | 8.792 KB |
|  microdados2018_arq6 | Planilha OpenOffice.org XML 1.0 | 1.540 KB | Não | 8.257 KB |
|  microdados2018_arq7 | Planilha OpenOffice.org XML 1.0 | 1.609 KB | Não | 8.608 KB |
|  microdados2018_arq8 | Planilha OpenOffice.org XML 1.0 | 1.606 KB | Não | 8.608 KB |
|  microdados2018_arq9 | Planilha OpenOffice.org XML 1.0 | 1.618 KB | Não | 8.608 KB |
|  microdados2018_arq10 | Planilha OpenOffice.org XML 1.0 | 1.606 KB | Não | 8.608 KB |
|  microdados2018_arq11 | Planilha OpenOffice.org XML 1.0 | 1.606 KB | Não | 8.608 KB |
|  microdados2018_arq12 | Planilha OpenOffice.org XML 1.0 | 1.604 KB | Não | 8.608 KB |
|  microdados2018_arq13 | Planilha OpenOffice.org XML 1.0 | 1.617 KB | Não | 8.608 KB |
|  microdados2018_arq14 | Planilha OpenOffice.org XML 1.0 | 1.604 KB | Não | 8.608 KB |
|  microdados2018_arq15 | Planilha OpenOffice.org XML 1.0 | 1.601 KB | Não | 8.608 KB |
|  microdados2018_arq16 | Planilha OpenOffice.org XML 1.0 | 1.605 KB | Não | 8.608 KB |
|  microdados2018_arq17 | Planilha OpenOffice.org XML 1.0 | 1.532 KB | Não | 8.608 KB |
|  microdados2018_arq18 | Planilha OpenOffice.org XML 1.0 | 1.613 KB | Não | 8.608 KB |
|  microdados2018_arq19 | Planilha OpenOffice.org XML 1.0 | 1.608 KB | Não | 8.608 KB |
|  microdados2018_arq20 | Planilha OpenOffice.org XML 1.0 | 1.616 KB | Não | 8.608 KB |
|  microdados2018_arq21 | Planilha OpenOffice.org XML 1.0 | 1.613 KB | Não | 8.608 KB |
|  microdados2018_arq22 | Planilha OpenOffice.org XML 1.0 | 1.323 KB | Não | 8.134 KB |

ENADE

- Uma análise desses microdados feita pelo jornal Folha de S.Paulo foi **premiada pelo próprio INEP** em 2018
- Depoimento de um dos jornalistas responsáveis pela análise dos dados:

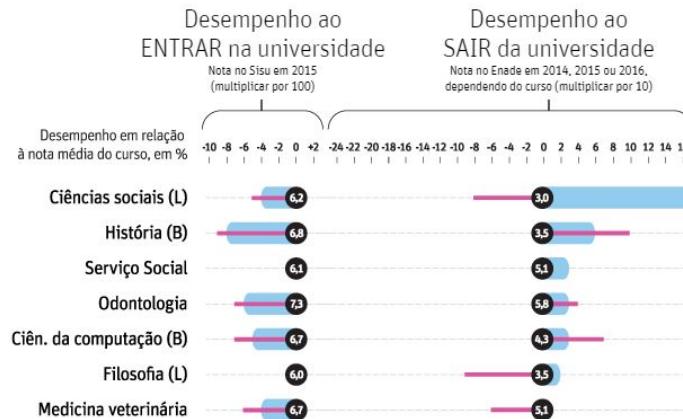
“hoje, os dados estão basicamente inúteis”

educação

Cotista tem nota boa na universidade, mas recua em cursos com matemática

★

O DESEMPENHO DOS COTISTAS EM RELAÇÃO AOS NÃO-COTISTAS



Salas seguras

Solução, sim; mas para **outro problema**

- **Permite acessar dados *identificados*** para pesquisas de interesse público
- Impõe condições legítimas para essa finalidade, mas restringe todas as pesquisas que não precisariam de dados identificados:
 - Cria incentivo para mais gente querer acessar dados restritos
 - Diminui o acesso
 - Acaba com a análise exploratória

Em 2023, 13 processos de acesso foram abertos e 10 deferidos.

Quantos *downloads* dos microdados havia antes?

Quantas pesquisas se beneficiavam?

Relação divulgada pelo Inep em 25.10.2023;

última atualização em 5.6.2023.

Em 2022 foram ~25 deferidos

Quem tem acesso aos dados hoje?

O GLOBO | Educação

Política | Brasil | Rio | Mundo | Economia | Saúde | Cultura | Esportes | Colunistas | Opinião | Clube | Newsletters

Brasil / Educação

Enem: saiba quais são as escolas com as 50 maiores notas do país e as top 10 por estado

Dados foram obtidos com exclusividade por GLOBO com pesquisadores da edtech AIO Educação, que retiraram informação diretamente da sala segura do Inep

Por Bianca Gomes e Bruno Alfano — Rio e São Paulo

16/10/2023 04h31 · Atualizado 18/10/2023

Questões

que merecem
aprofundamento

- O que a sociedade brasileira deixou de saber sobre os impactos da desigualdade na educação após a medida?
- Quantas pessoas deixaram de ter acesso aos dados e qual era o valor público que eles geravam para a sociedade?
- Qual o tamanho do gargalo da sala segura?
- Quantas empresas privadas lucram com dados que são restritos à sociedade civil?
- Por que os dados sobre docentes foram completamente deixados de lado nos ex-microdados dos censos?

Obrigada!

*Aponte o celular para o
QR Code ao lado para
conhecer mais sobre
o nosso trabalho.*

