

Eixo 2: Pesquisa em Classificação, Catalogação e Indexação

## **VISUALIZAÇÃO E ANÁLISE DE DADOS BIBLIOGRÁFICOS DAS PUBLICAÇÕES SERIADAS NO BRASIL**

### *Visualization and analysis of bibliographic data of serial publications in Brazil*

#### **Bruno Costa**

Doutor

Instituto Federal de Educação, Ciência e Tecnologia do Rio de Janeiro (IFRJ), Rio de Janeiro, Brasil.

Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT), Brasília, Distrito Federal, Brasil.

ORCID: <https://orcid.org/0000-0001-9655-8698>

#### **Ana Carolina Simionato Arakaki**

Pós Doutorado em Ciência da Computação – Universidade de São Paulo (USP), São Paulo, Brasil. Doutorado em Ciência da Informação – Universidade Estadual Paulista (UNESP), Marília, São Paulo, Brasil. Professora Adjunta - Universidade Federal de São Carlos (UFSCar), São Carlos, São Paulo, Brasil. Coordenadora de Serviços Bibliográficos no Instituto Brasileiro de Informação em Ciência e Tecnologia (COBIB/IBICT), Brasília, Brasil.

E-mail: [anaarakaki@ibict.br](mailto:anaarakaki@ibict.br)

ORCID: <http://orcid.org/0000-0002-0140-9110>.

#### **Renan Barbosa**

Ensino Médio

Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT), Brasília, Distrito Federal, Brasil.

#### **Renan Luiz**

Ensino Médio

Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT), Brasília, Distrito Federal, Brasil.

## Resumo

A análise estatística dos dados de publicações seriadas impressas no Brasil é dificultada pela dispersão, o formato físico e a falta de padronização dessas informações. Este estudo propõe um dashboard que facilita a visualização e a análise estatística dos dados relacionados às publicações seriadas, tornando o processo mais simples e intuitivo. A solução visa melhorar a interpretação dos dados e aumentar a acessibilidade às informações, promovendo avanços significativos na análise dentro desse contexto.

**Palavras-Chave:** Dashboard; Análise de dados; Visualização de dados.

## Abstract

The statistical analysis of data related to printed serial publications in Brazil is hindered by the dispersion, physical format, and lack of standardization of these records. This study proposes a dashboard that facilitates the visualization and statistical analysis of data related to serial publications, making the process simpler and more intuitive. The solution aims to enhance data interpretation and increase accessibility to this information, thereby promoting significant advances in analysis within this context.

**Keywords:** Dashboard; Data Analysis; Data visualization.

## 1 INTRODUÇÃO

A catalogação e a localização de publicações seriadas sempre desempenharam um papel fundamental no apoio à pesquisa, especialmente para pesquisadores. Com o avanço tecnológico, muitas dessas publicações passaram a ser disponibilizadas em formato digital, facilitando o acesso. Entretanto, antes da digitalização, o principal desafio era identificar a localização física de revistas nas diversas bibliotecas nacionais, tarefa para a qual o Catálogo Coletivo Nacional de Publicações Seriadas (CCN), sob a coordenação do Instituto Nacional de Informação em Ciência e Tecnologia (Ibict), foi desenvolvido. Até os dias de hoje, o

CCN continua a ser uma ferramenta essencial, especialmente porque muitas revistas, particularmente em áreas como Medicina, Geologia e Direito, publicadas antes de 1990, não estão digitalizadas.

No Brasil, o CCN consolidou-se como a maior base de registros de publicações seriadas impressas, abrangendo tanto títulos nacionais quanto estrangeiros. O CCN mantém em seu banco de dados informações detalhadas sobre mais de 62 mil títulos de periódicos e mais de 3 milhões de transcrições de coleções provenientes de mais de 600 bibliotecas brasileiras, de instituições de ensino e pesquisa distribuídas por todas as regiões do país. No entanto, a extração de dados estatísticos a partir deste vasto repositório para fins de pesquisa, otimização e aprimoramento do serviço tem se mostrado desafiadora devido à ausência de ferramentas adequadas de análise e visualização de dados.

O presente estudo visa explorar as variáveis relacionadas à base de dados do CCN, indo além da simples identificação de conexões, ao buscar apresentá-las de maneira clara e acessível aos usuários. Para isso, foi desenvolvido um dashboard<sup>1</sup> como ferramenta de visualização, projetado para traduzir os resultados da análise em representações gráficas intuitivas, facilitando a compreensão das correlações descobertas. O dashboard desenvolvido tem por objetivo não apenas enriquecer o conhecimento atual sobre as publicações seriadas impressas, mas também fornecer uma base sólida para o desenvolvimento contínuo e aprimorado do CCN, promovendo avanços significativos na análise e interpretação de dados.

## 2 DESENVOLVIMENTO

A extração de informações estatísticas acerca de publicações seriadas enfrenta uma série de desafios, que vão desde questões relacionadas aos custos e à padronização, até a ausência de informações e a dificuldade de acesso. Reconhecendo essas dificuldades, foi desenvolvida uma solução de visualização de

---

<sup>1</sup> O Dashboard CCN está disponível pelo endereço [https://lookerstudio.google.com/u/0/reporting/8c58534d-22fe-4937-9cdd-d6741878a68f/page/p\\_7uu2pe22ad](https://lookerstudio.google.com/u/0/reporting/8c58534d-22fe-4937-9cdd-d6741878a68f/page/p_7uu2pe22ad)

dados especificamente voltada para o Catálogo Coletivo Nacional de Publicações Seriadas (CCN). Essa solução visa facilitar o acesso às publicações seriadas impressas e possibilitar a análise de dados, bem como a apresentação de resultados de estudos já realizados com base nesse acervo.

Além disso, um dos principais objetivos foi criar uma interface de visualização que fosse simples, intuitiva e objetiva, tornando as informações mais acessíveis aos usuários. Para alcançar isso, foram aplicadas técnicas de storytelling (Nussbaumer, 2019), utilizando escolhas visuais eficazes, paletas de cores adequadas e uma distribuição espacial harmoniosa dos elementos, visando otimizar a experiência do usuário e a interpretação dos dados.

Com a criação de visualizações de dados gerais do CCN, passou-se a realizar estudos detalhados sobre esses dados. Um desses estudos focou em identificar correlações não evidentes entre as variáveis disponíveis. O objetivo foi explorar como essas variáveis podem, direta ou indiretamente, influenciar e explicar o comportamento de outras variáveis dentro do mesmo conjunto de dados. Ao adotar essa abordagem analítica, busca-se não apenas descobrir padrões explícitos, mas também desvendar interações sutis que possam proporcionar uma compreensão mais completa e aprofundada dos dados.

### 3 PROCEDIMENTOS METODOLÓGICOS

A criação do dashboard foi realizada utilizando o *Looker Studio*, uma ferramenta *online* do Google para a criação de relatórios e painéis informativos personalizados. Através do uso de técnicas de *storytelling* (Nussbaumer, 2019), buscou-se tornar a análise dos dados intuitiva, simples e objetiva. O estudo foi direcionado para uma visão quantitativa, limitando-se a apresentar as principais distribuições entre as variáveis que compõem o banco de dados do CCN, como assunto, frequência de publicação e área do conhecimento. Considerando que o volume de dados não era extremamente grande, o que poderia demandar técnicas

mais avançadas de coleta e tratamento, foram utilizados todos os registros existentes no banco de dados.

No estudo das correlações não evidentes, a pesquisa adotou uma abordagem quantitativa descritiva-exploratória para explorar as possíveis associações entre as variáveis do banco de dados do CCN. Todas as análises foram realizadas utilizando o software estatístico R, garantindo precisão e robustez nos resultados.

Para verificar a existência de relações significativas entre as variáveis, foi aplicado o Teste Qui-Quadrado, uma técnica estatística utilizada para avaliar se a distribuição observada das frequências em diferentes categorias difere significativamente da distribuição esperada. A quantidade e o nível de detalhe das informações fornecidas pelo Teste Qui-Quadrado tornam-no uma ferramenta valiosa para a análise de dados nominais (McHugh, 2013). Esse teste é particularmente eficaz para identificar associações entre variáveis categóricas, permitindo a descoberta de correlações não evidentes.

Além disso, para quantificar a força dessas associações, foi utilizado o Teste de Cramer. O Teste V de Cramer é uma medida estatística que mede a força da associação entre variáveis categóricas, com valores que variam de 0 a 1, onde valores mais próximos de 1 indicam uma associação mais forte entre as variáveis (Cramér, 1946). Este teste é uma medida de associação que complementa o Teste Qui-Quadrado. A análise foi essencial para entender a intensidade das correlações identificadas e destacar aquelas de maior relevância no contexto da pesquisa. Por fim, os resultados foram integrados ao dashboard, permitindo aos usuários testar na prática as principais correlações identificadas.

## 4 RESULTADOS

O principal resultado deste trabalho é a criação de um dashboard que centraliza os dados mais relevantes do CCN de forma simples e intuitiva, utilizando técnicas de storytelling (Nussbaumer, 2019) para facilitar a interpretação dos dados. Além disso, o dashboard integra os resultados dos estudos realizados com base

nesses dados, proporcionando uma ferramenta robusta para análise e visualização. A seguir, os resultados serão apresentados em duas grandes frentes, detalhando as funcionalidades e os insights obtidos a partir do painel.

#### 4.1 VISUALIZAÇÃO DOS DADOS QUE COMPÕEM O CCN

O primeiro painel reúne informações gerais acerca das publicações seriadas disponíveis no CCN. Este painel apresenta gráficos que facilitam a compreensão das informações mais relevantes. A seguir, são destacados os principais gráficos exibidos no dashboard. Tendo em vista a técnica de data storytelling, nos próprios painéis, são incluídas informações necessárias para promover a compreensão dos dados.

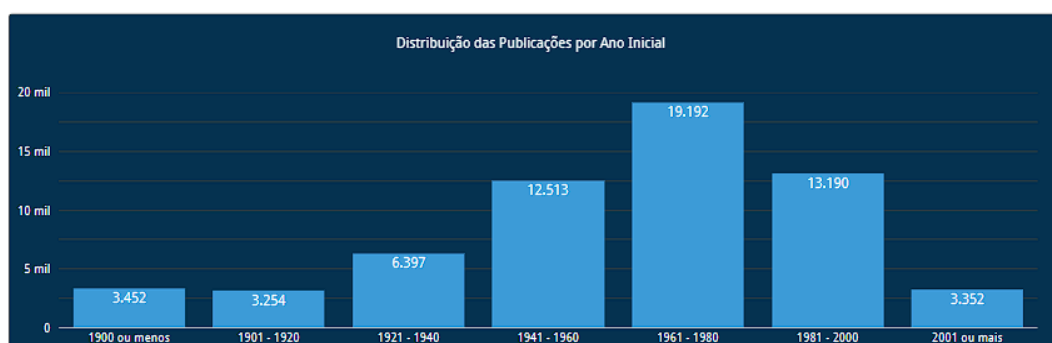
**Figura 1** – Seção inicial do *dashboard* CCN



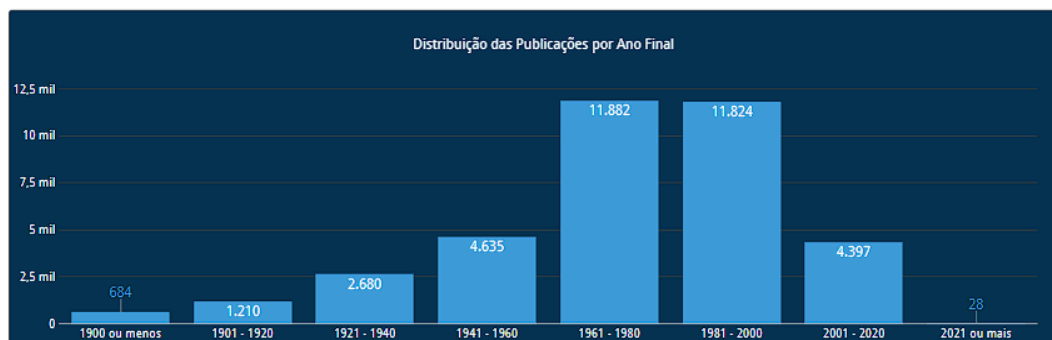
**Fonte:** <https://lookerstudio.google.com/s/rvs79HiWP7M>

**Figura 2 – Publicações seriadas pelo seu ano inicial e final**

Em questão de conteúdo, a maioria é do tipo Técnico e/ou Científico

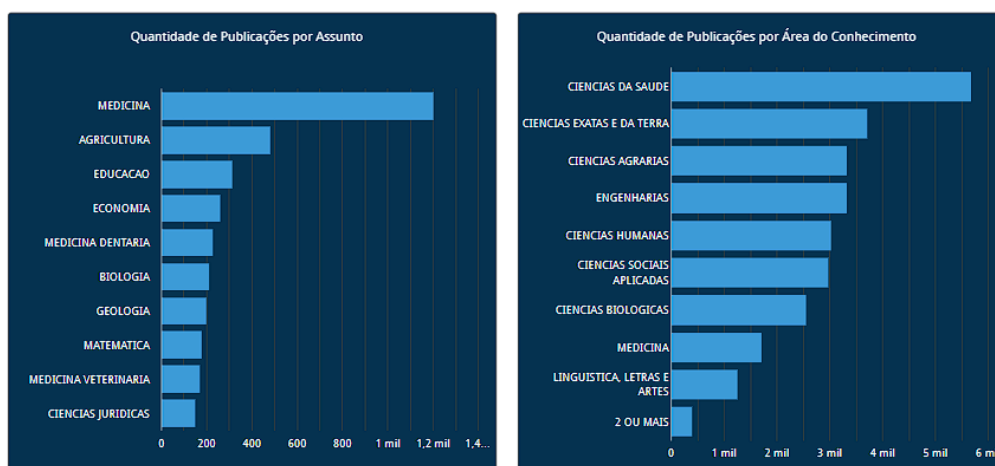


Grande parte das publicações se iniciaram no século passado. Além disso, há um grande volume de publicações que se enceraram nos últimos 3 (três) anos.



**Figura 3 – Publicações seriadas por assunto e área do conhecimento**

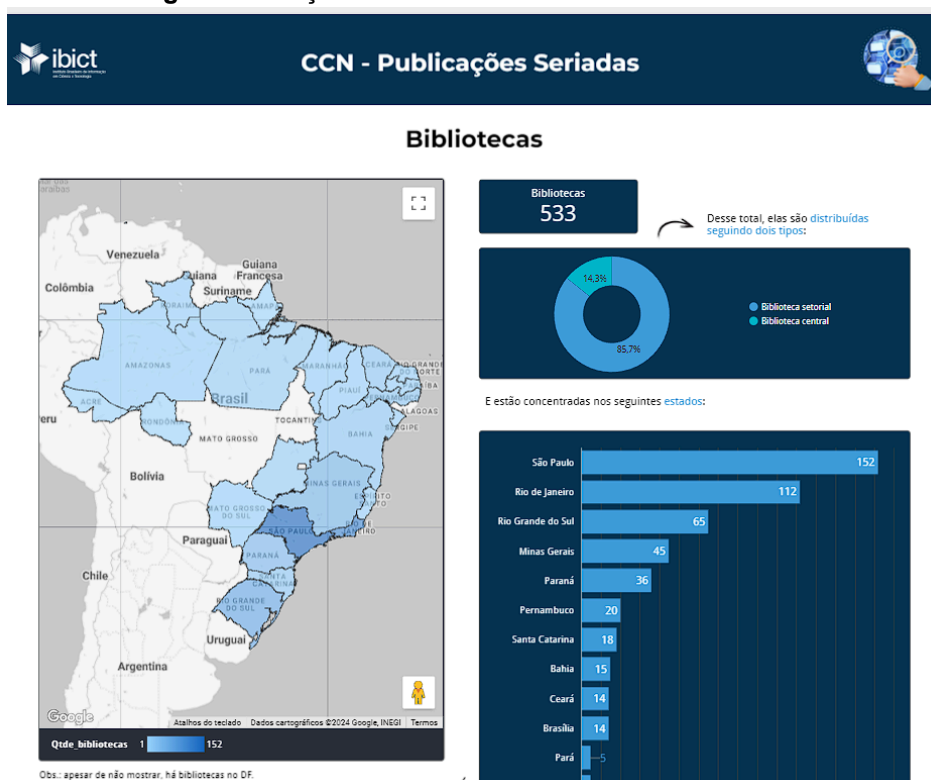
Quando falamos de assunto e área do conhecimento, temos as seguintes distribuições:



Vale ressaltar que além de demonstrar por meio de gráficos e cartões dados sobre as publicações seriadas, há uma seção exclusiva para tratar sobre os dados

das bibliotecas cadastradas no CCN, o qual é possível visualizar de maneira parcial na imagem abaixo.

**Figura 4 – Seção “CCN - Biblioteca” do Dashboard CCN**



## 4.2 CORRELAÇÕES NÃO EVIDENTES

Para a realização do estudo das correlações não evidentes, foram testadas as seguintes variáveis do banco de dados do CCN: assunto, área de conhecimento, frequência, situação (que diz se uma dada publicação seriada está suspensa ou ativa), conteúdo, ano de publicação, editora, bibliotecas relacionadas e software utilizado por essas bibliotecas, obtendo os seguintes resultados relevantes quanto à relação de variáveis. A Figura 5 apresenta alguns resultados da correlação entre tais variáveis; a Figura 6 exibe a tela de filtragem das variáveis; a Figura 7 apresenta a visualização gráfica da correlação.

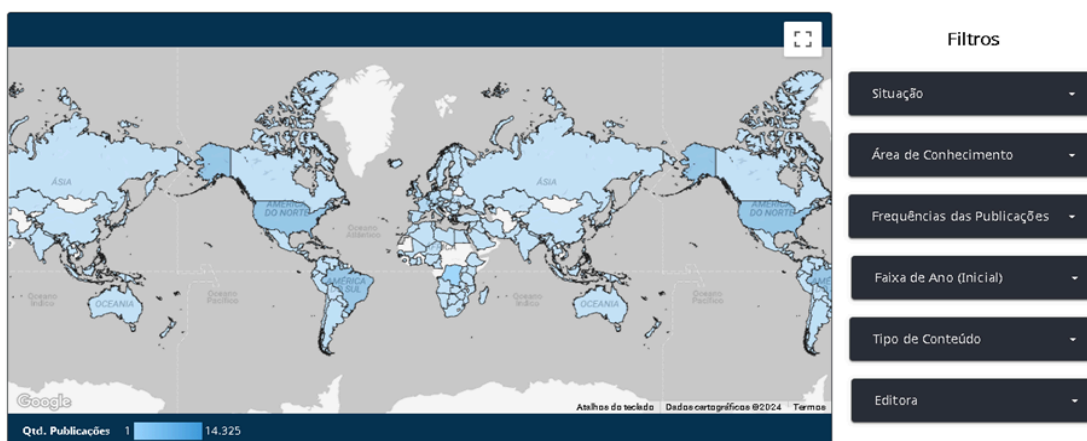


**Figura 5 – Resumo dos resultados para cada variável**

Teste	p-valor	Resultado	Cramer	Classificação
Assunto vs Conteúdo	<0.0001	Rejeita a hipótese nula	0.5276669	Associação moderada forte
Área de Conhecimento vs Editora	<0.0001	Rejeita a hipótese nula	0.8217522	Associação muito forte
País vs Editora	<0.0001	Rejeita a hipótese nula	0.9378392	Associação muito forte
Frequência vs Conteúdo	<0.0001	Rejeita a hipótese nula	0.5087201	Associação moderada forte
Frequência vs Editora	<0.0001	Rejeita a hipótese nula	0.7445655	Associação forte
Situação vs Editora	<0.0001	Rejeita a hipótese nula	0.8069543	Associação muito forte
Situação vs Ano	<0.0001	Rejeita a hipótese nula	0.40981	Associação moderada forte
Conteúdo vs Editora	<0.0001	Rejeita a hipótese nula	0.722845	Associação forte
Editora vs Ano	<0.0001	Rejeita a hipótese nula	0.7957987	Associação forte

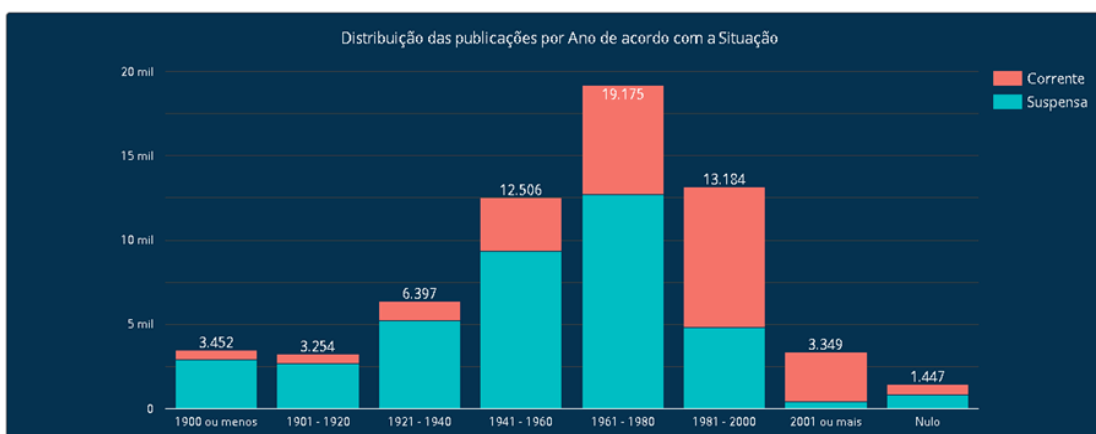
**Figura 6 – Seção “CCN - Correlações Não Evidentes” do Dashboard CCN**

Publicações são representadas geograficamente, enquanto variáveis como editora, frequência e publicações correntes servem como filtros interativos para o mapa e para a página inteira, proporcionando uma visão dinâmica e personalizada das conexões globais e conceituais dentro do CCN



**Figura 7 – Publicações seriadas por ano inicial e situação**

A seguir, foi confrontado o ano de publicação com a situação atual das publicações, a correlação aparenta ser bem clara, até 1980 temos a maioria das publicações nos dias atuais estão suspensas, de 1981 para frente a situação se inverte, passando a ter mais publicações correntes.



## 5 CONCLUSÃO

Os objetivos propostos no estudo foram integralmente atingidos com a criação e disponibilização de um dashboard que centraliza os principais dados do Catálogo Coletivo Nacional de Publicações Seriadas (CCN) e os resultados de análises aplicadas a esse banco de dados. A ferramenta visual e interativa proposta representa um avanço significativo na capacidade de exploração e análise dos dados, proporcionando aos pesquisadores uma plataforma robusta para a identificação de padrões e para a tomada de decisões embasadas. Tal ferramenta se revela crucial não apenas para o enriquecimento de pesquisas acadêmicas, mas também para consultas da comunidade civil e como suporte a projetos de extensão e pesquisa no ambiente universitário.

A apresentação dos resultados por meio do dashboard confere um caráter acessível e ao mesmo tempo impactante às descobertas, democratizando o acesso à informação e facilitando a compreensão das dinâmicas dos dados analisados por um público mais amplo, que vai além dos especialistas na área. Essa acessibilidade é fundamental para expandir o uso e a compreensão dos dados do CCN, contribuindo para uma análise mais eficiente e abrangente das publicações seriadas impressas no Brasil.

Ademais, a solução desenvolvida abre perspectivas para futuros enriquecimentos, como a integração de dados externos e internos do CCN, ou até de outros serviços disponibilizados pelo Ibict. Com a ferramenta, é possível identificar padrões, compreender volumetrias e comportamentos, tanto atuais quanto históricos, e detectar variáveis com correlações estatisticamente significativas. Isso posiciona o dashboard como uma ferramenta essencial para aprofundar e orientar pesquisas não apenas na área da Biblioteconomia, mas também em áreas específicas como a Medicina, dada a amplitude e diversidade do acervo.

Em síntese, o dashboard não apenas expande o conhecimento existente, mas também estabelece uma base sólida para o desenvolvimento contínuo e

aprimorado do CCN, promovendo avanços substanciais na análise e interpretação dos dados relacionados às publicações seriadas impressas no Brasil.

## 6 REFERÊNCIAS

MCHUGH, Mary L. The Chi-square test of independence. **Biochemia Medica**, v. 23, n. 2, p. 143-149, 2013. DOI: 10.11613/BM.2013.018.

NUSSBAUMER, Cole. **Storytelling com dados: um guia sobre visualização de dados para profissionais de negócios**. 2. ed. Rio de Janeiro: Alta Books, 2019. 242 p.

CRAMÉR, Harald. **Mathematical Methods of Statistics**. Princeton: Princeton University Press, 1946. ISBN 9780691080048.