

3-2. データの引用

データを引用することの大切さ

研究成果を発表する際にはその情報源を引用しなければならない。引用文献リストは、著者の主張、理論、考察の信頼性を担保する上で極めて重要。

データセットにおいても引用は文献と同様に重要。しかし、適切なデータ引用の実践の進展は遅く、必要な情報は引用文献リストにない。

適切なデータセットの作成者は、他の研究者の研究に貢献するだろうが、データセットの功績が認められることは少ない。

2014年に、FORCE11が適切なデータ引用に関する共同宣言を発表し、多くの学術団体や研究者が支持。

メモ

データ引用原則に関する共同宣言 FORCE11

1. 重要性：出版物と同様に引用されるべき
2. 功績と帰属：貢献度に応じて適切な帰属と功績を認める
3. 証拠：データに基づいて主張する場合は引用すべき
4. ユニークな識別子：世界的に利用され、ユニークな識別子を付与
5. アクセス性：データや関連するメタデータ等のアクセスを進める
6. 永続性：識別子とメタデータの組合せは永続する
7. 特殊性と検証可能性：引用情報が元のデータと同じであることを保証する、十分な起源と変わらない情報を含む
8. 相互運用性と柔軟性：データ引用は分野により柔軟であるべきだが分野間の相互運用性は保つ

<https://www.force11.org/group/joint-declaration-data-citation-principles-final> より

メモ

データ引用の方法

研究で参照したデータセットは、他の利用者がアクセスできるよう適切な情報を含まなければならない。

DataCite は、以下の引用要素を推奨

Creator (PublicationYear): Title. Publisher. Identifier

Identifier は、DOI や URL 等の一意に識別するもの。

可能ならば Version と ResourceType も追加する。

Creator (PublicationYear): Title. Version. Publisher. ResourceType. Identifier

UK Data Serviceは、主題を表すタイトルに加えて、データがカバーする地理的情報と時間的範囲を示すことを推奨。

メモ

データの引用情報の例

Richardson, Elizabeth A. (2009). Carstairs deprivation scores for Scotland by CATT2, 1981, 1991, 2001 [Dataset]. University of Edinburgh. School of GeoSciences. <https://doi.org/10.7488/ds/155>

作成者、作成日、データタイトル、識別子

引用文献と同じスタイルや順番を採用する。文献管理ソフトが助けとなる。

継続的に更新される動的なデータベースの場合、利用されたデータベースのバージョンを正確に参照することはほぼ不可能。この場合はダウンロード日を含める。

生命科学では、時系列データを参照するとき、単純にデータベース名とアクセス番号を提供する。

European Nucleotide Archive accession number CY115901, or NCBI SNP database accession number ss472331023

メモ

研究データに関する情報を管理するには

1. ファイル構成を見ただけで、データの内容を予測できるよう規則性を維持する
2. データを再利用できるように、その収集方法や分析方法、内容について、第三者が十分理解できるように文書を残す
3. インターネットで共有することを考慮して、十分なメタデータを付与する
4. データを利用した場合は引用情報を提示する

メモ