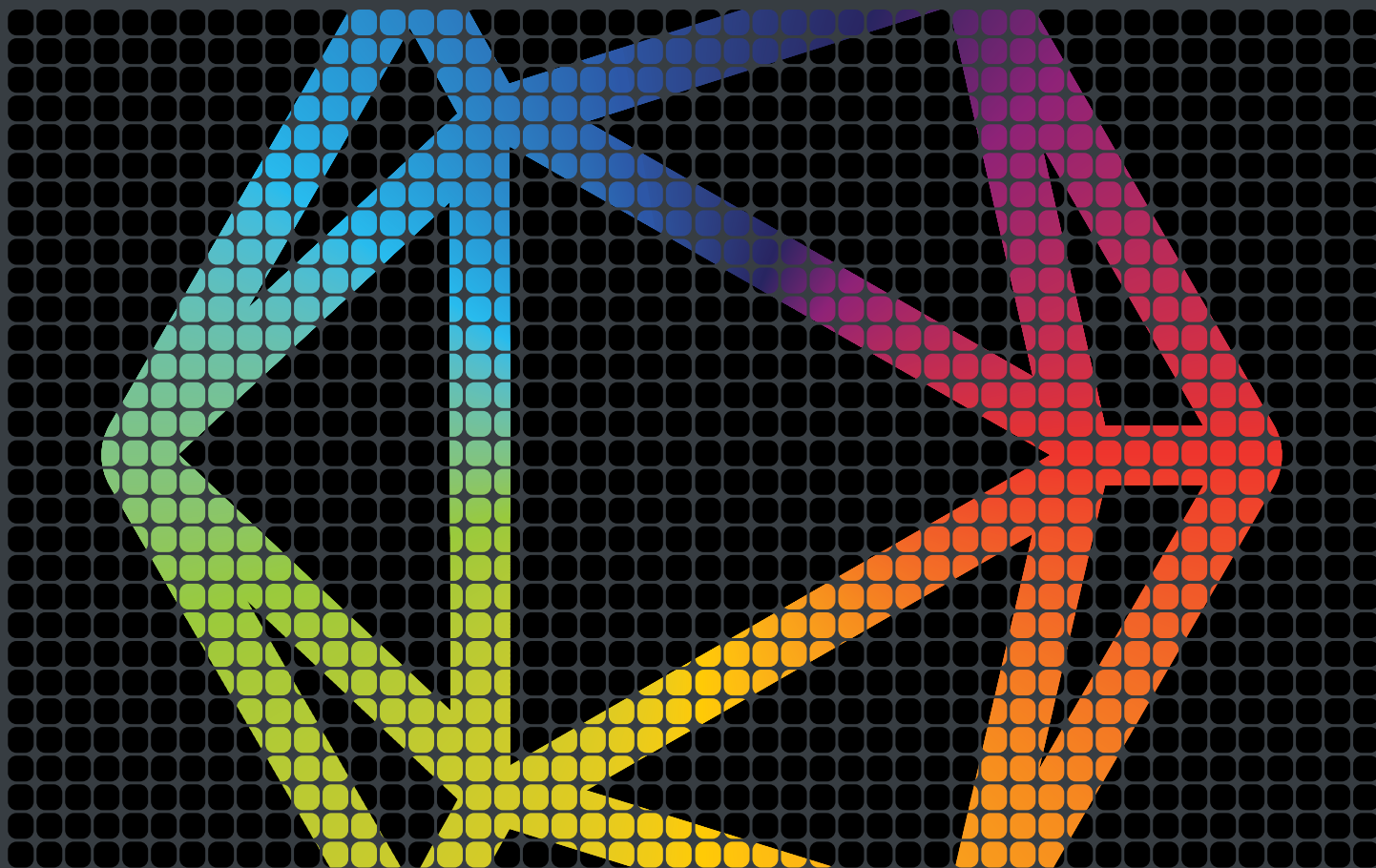




Dimensions Report

Dimensionsデータガイド

学術研究の全容を把握するための
インフラストラクチャ開発のアプローチ

Christian Bode, Christian Herzog, Daniel Hook & Robert McGrath

APRIL 2019

About Dimensions

Dimensions®は近代的で革新的な、相互リンクされた研究情報のインフラストラクチャです。出版物や助成金、臨床試験、特許、政策ドキュメントなどを一括で検索し、アクセスを容易にする新しいツールです。Digital Science 傘下のポートフォリオ企業のお客様やパートナーからのフィードバックを受けて、Dimensionsの開発はスタートしました。6つのポートフォリオ（ReadCube、Altmetric、Figshare、Symplectic、ÜberResearch、DS Consulting）のダイナミックなコラボレーションの結果、Dimensionsが誕生しました。研究サイクルにおけるさまざまな問題点に取り組み、研究エコシステムの多くのステークホルダーに関わってきたこれらのチームは、イノベーションへの真の情熱を共有し、独自の経験と意見と価値をもってDimensionsの開発に貢献しました。詳しくは www.dimensions.ai をご覧ください。

About Digital Science

Digital Science®は研究サイクルの各段階で科学研究コミュニティのニーズに応えるテクノロジー・カンパニーです。私たちはすべての研究プロセスをもっとオープンで効率的・効果的なものにするために、革新的なビジネスとテクノロジーに投資し、育成・サポートしています。私たちは皆さんと共に、研究のあり方を変えられることができると信じています。詳しくは www.digital-science.com をご覧ください。

Acknowledgements

このレポートを作成するために時間をかけて必要なデータを提供してくれた開発チーム全員に謝意を評します。

このレポートは Holtzbrinck Publishing Group 傘下の Digital Science 社によって提供されています。

Dimensionsについてのご質問は、メールにて info@dimensions.ai までお寄せください。Digital Scienceへのご連絡は info@digital-science.com もしくは 625 Massachusetts Avenue, Cambridge, MA, 02139 USA までお寄せください。



Contents

1. 研究の全体像をリンクする	2
2. すべてをリンクして、ユーザーのために強化	3
全文テキストを検索可能	4
機械学習にもとづいた研究トピック分類	4
Dimensionsのカテゴリ：オーストラリア・ニュージーランド標準研究分類（ANZSRC）	4
その他の分類システム	5
GRID：研究機関の名寄せ	6
人物同定のチャレンジ：出版物から助成金、特許、臨床試験まで	7
引用と謝辞による研究文脈の追加	7
3. 出版物を超えて： 可能な限り多くのソースからコンテンツを収録	8
Dimensionsは他のデータベース、例えばGoogle ScholarやPubMed、Scopus、Web of Science等とどう違うのか？	9
さまざまなデータベースの引用カウント	9
現在の収録範囲と品質を出発点として	10
4. 確かな未来の兆候としての助成金情報	11
Dimensionsの助成金データ	12
5. 出版物と引用情報	14
Dimensionsの出版物データと引用情報：データベースは「審査基準」ではありません	14
品質保証のフィルタリングとしてのホワイトリスト/ブラックリスト	15
Dimensionsの出版情報と引用データの統合	15
アカデミックな注目を超えて：Dimensionsのオルトメトリクス	16
オープンアクセス、オープンな引用データ、そしてDimensions	17
Dimensionsの出版・引用データ	17
6. 臨床試験情報：臨床応用に向けた研究成果	19
Dimensionsの臨床試験情報	19
7. 特許情報：実用・商用利用に至った研究成果	21
Dimensionsの特許情報	21
8. 政策ドキュメント：政策やガイドラインに影響を与えた研究	23
Dimensionsの政策ドキュメント情報	23



研究の全体像を リンクする

Digital Scienceが
100以上の
パートナーと共に
開発した
Dimensions

出版物情報
と引用データ
の無料提供

研究コミュニティ
の支援と
エンパワーメント

ご意見・ご要望を
お寄せください

より良いサービス
を目指して

Dimensionsは、Digital Science社とその開発パートナーが抱える2つの大きな制約に対処するために考案されました。一つ目の制約は、既存のソリューションが研究の全体像を出版物とその引用情報だけで理解しようとしていたことです。二つ目の制約は、それらの既存のソリューションがどんなデータを収録しているのかわかりにくかった、ということです。多くの出版情報リンクは有料のアプリケーションに閉じ込められ、その利用は制限されていましたし、機能的なAPIもありませんでした。多くの商用データは不完全で、肝心な利用目的に耐えないこともありました。

これらの制約を克服して研究支援のイノベーションを活性化すべく、私たちは100以上の開発パートナー（研究機関や助成機関）と密接な連携をはかり、ついに研究プロセスの全体をカバーする統合データベースを実現しました。助成金獲得から研究へ、そしてその成果物に対する注目度や、学術的・商業的な応用、政策決定に与えた影響まで、あらゆる側面を一貫してリンクしたのです。

Dimensionsのビジョンの中核として、私たちは研究に革新をもたらすことを望んでいました。その重要な一部となったのが、Dimensionsの無料版 (<https://app.dimensions.ai>) での出版物情報と引用データの提供、またオープンなDimensions Metrics APIとDimensions Badges (<https://badge.dimensions.ai>)によるメトリクスの提供です。どちらも非商用利用に限って無料で使えます。

学術コミュニティへのもう一つの支援としては、エンパワーメントがあります。現在の研究評価の潮流では、評価の対象となる大量の文献に対処するためにメトリクスの利用が奨励されています。よりオープンな出版物とオープンなデータソース、よりオープンな評価の枠組みとオープンなメトリクスが必要とされているのは明らかです。研究のベストな成果と文脈を語るのに必要なメトリクスの開発を、学術コミュニティ自体が司るためのシステム、それがDimensionsです。

本ドキュメントはDimensionsの収録内容の概要について解説しています。Dimensionsの収録範囲や対象があなたのニーズに即しているかどうか、もしご要望があればぜひお知らせください。

Dimensionsの開発にあたって最も重要なのは、私たちがこれを研究コミュニティと共にさらに発展させようとしているということです。どのようなご意見も歓迎します。info@dimensions.ai までご連絡ください。



Dimensionsの収録データ概要	
コンテンツのタイプ	レコード件数
出版物 (Publications)	1億件
助成金 (Grants)	4,600万件
特許 (Patents)	3,800万件
臨床試験 (Clinical Trials)	45万5,000件
政策ドキュメント (Policy Documents)	42万2,000件
オルトメトリクス指標のあるレコード	1,000万件
総データ件数	1億5,300万件

すべてをリンクして、 ユーザーのために強化

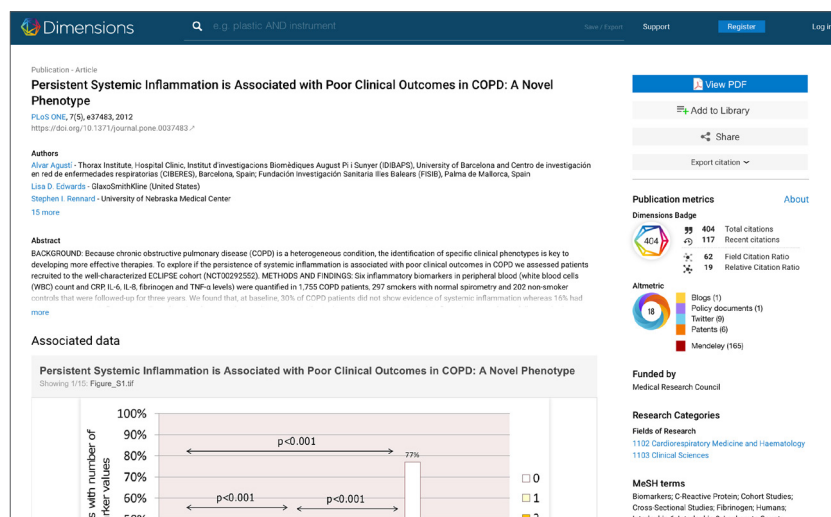
複数のソースから収集して相互にリンクし統合されたデータ、それがDimensionsです。製品のスコープと方向性を開発パートナーと検討するにあたって、これこそが最も重要な機能でした。統合された全体像から新たな知見が得られるということに、皆さんが同意されました。以下に、主なアプローチと概要を示します。

機械学習と人工知能をベースとしたアプローチにより、自動的に情報を抽出して関連性を見出し、相互リンクを実現します。コンテンツ抽出と強化のプロセスは可能な限り自動化することで、研究者への出版・引用データの無料提供と、契約機関には現実的な価格帯を実現します。自動化のプロセスはよりオープンで低コストを実現しますが、データに関しては継続して改善していかなければならない点もあります。もしおかしなデータがあれば、ぜひお知らせください。私たちは常にプロセスを改善し、Dimensionsが提供するデータとサービスの向上に努めています。ご連絡は support@dimensions.ai までお願いします。

以下に示すのは、無料版Dimensionsの論文レコードの例です。他のコンテンツにリンクして、関連情報を参照することができます。

重要なのは
さまざまなデータを
相互リンクすること

自動化による
効率的・効果的な
プロセスと、
さらなる高品質を
目指して



サンプルデータ :
Persistent Systemic Inflammation is
Associated with Poor Clinical Outcomes
in COPD: A Novel Phenotype" (DOI
10.1371/journal.pone.0037483)

Publication references	44
Associated data sets	15
Supporting Grants	2
Publication citations	401
Patent citations	5
Linked Clinical trials	2
Policy document citations	1
Altmetric Attention Score	18



全文テキスト 検索で関連情報 を見逃さない

論文レベルの 指標は 論文レベルの 分類から

自然言語検索と 機械学習で 内容に即した 分類を

オーストラリアと ニュージーランドで 開発された 標準分類

全文テキストを検索可能

Dimensionsの無料版というディスカバリーサービスを研究者に提供するにあたって、出版情報の発見可能性を高めるために、私たちは全文テキスト検索のアプローチをとりました。Dimensionsでは6,900万件以上の出版物の全文テキストが検索可能です。例えばタイトルと抄録だけを対象にCRISPRと検索すると、15,000件がヒットしますが、全文テキストを対象にすれば77,000件以上が検索できます。フィルタリング機能を使って検索結果を絞り込むことで、Dimensionsの全文検索はパワフルなディスカバリーツールとなります。

機械学習にもとづいた研究トピック分類

Web of ScienceやScopusなど既存のデータベースでは、ドキュメントの分類は掲載ジャーナルにもとづいて行われるのが常でした。このアプローチは、データベースの収録や引用ベンチマークの方法など、さまざまな予期せぬ影響を与えることになりました。

技術は進歩し、自然言語処理や機械学習、人工知能などが近年目覚ましい発展を遂げています。Dimensionsはこうした技術を活用して、実際の課題にこれまでとは違うアプローチで臨みました。助成金情報や特許、臨床試験、政策ドキュメントなどを一貫して分類しようとするならば、ジャーナルの分類は役に立ちません。私たちが選んだのは、既存の分類システムと機械学習のアプローチを組み合わせ、ドキュメントのタイプに関わらず、一貫した分類を自動的に付与する方法でした。

分類アルゴリズムのトレーニングに使うことができるデータセットを備えた、すでに確立された研究分類システムが必要でした。私たちは、広く研究分野をカバーする分類システムとトレーニング用のコーパスを備えた、オーストラリアとニュージーランドで使われている研究分類システムを無料版Dimensionsの一部として採用したのです。

Dimensionsのカテゴリ：

オーストラリア・ニュージーランド標準研究分類(ANZSRC)

FOR (Fields of Research) はオーストラリア・ニュージーランド標準研究分類 (ANZSRC) の一部として2008年に開発されました。すべてのR&D活動を一貫して分類することができるANZSRCは、オーストラリアとニュージーランドの研究教育において広く利用されています。

FOR分類はDivision、Group、Fieldという三階層から構成されます。22のDivisionは広範な主題領域もしくは研究分野を表していて、157のGroupと1238のDivisionはそれぞれ詳細な下位分類を表します。Dimensionsでは、このうち第二階層 (Group) のみを採用しています。人手を使ってコーディングされた研究助成金情報をコーパスとして、やはり人手を使って付与するはずのFORコードをアルゴリズムを使って再現し、実際に付与されたコードと付き合わせては、またアルゴリズムを走らせるという、リバースエンジニアリングの手法をとりました。

FOR分類はハイレベルで学術研究の全ての領域をカバーしているので、主題領域を見渡すような粒度の粗い調査に有効です。このため、アカデミア全体の比較分析にもFORは適しているのです。



その他の分類システム

FORコードに加えて、その他の分類システムもDimensionsには実装されています。異なる分類スキームの選択は、主に研究助成機関からの要請で、その多くは生物医学に特化していました。これらの分類スキームの実装には、やはり前述のような機械学習のアプローチをとりました。以下にその例を挙げます。

- RCDC ([Research, Condition, and Disease Categorization](#)) は米国国立衛生研究所 (NIH) が連邦議会が要請する公式報告で採用している分類スキームです。ÜberResearchチームはNIHでのRCDC実装にあたって技術提供を行い、現在に至るまでそれをサポートしています。
- HRCS ([Health Research Classification System](#)) は英国のほぼすべての生物医学系の助成機関が保健・生物医学に特化したプロジェクトを分類するために採用しているシステムです。HRCSには、RAC (Research Activity Codes) と HC (Health Categories) の2種類があります。

米国RCDCや
英国HRCSを
分類スキームとして
実装

同様のアプローチで、他の分類システムについても比較的簡単に実装することができます。特定の主題分類が必要な機関向けに、いくつか追加のスキームがすでに実装されました。例えば、特定の国で利用されているシステムや、特定のトピックに的を絞ったシステムなどです。また、必要に応じて、Dimensionsがまだ収録していないドキュメントに分類を施すことも可能です。詳細については、Dimensionsチームまでご相談ください。

必要に応じて
新たな
分類システムも
導入可能



機関名の 表記ゆれを 解決するために

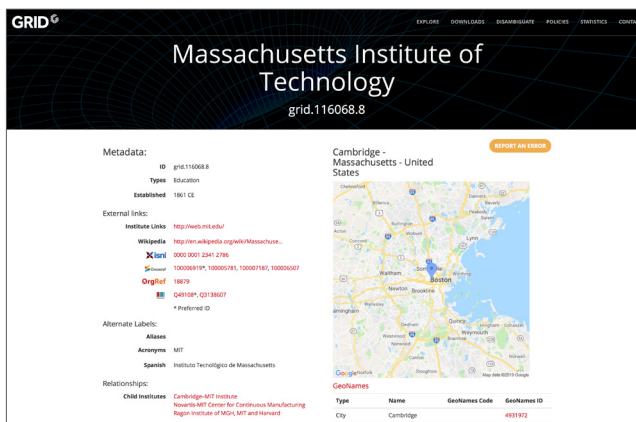
GRIDは Digital Scienceが 提供する オープンデータです

GRID：研究機関の名寄せ

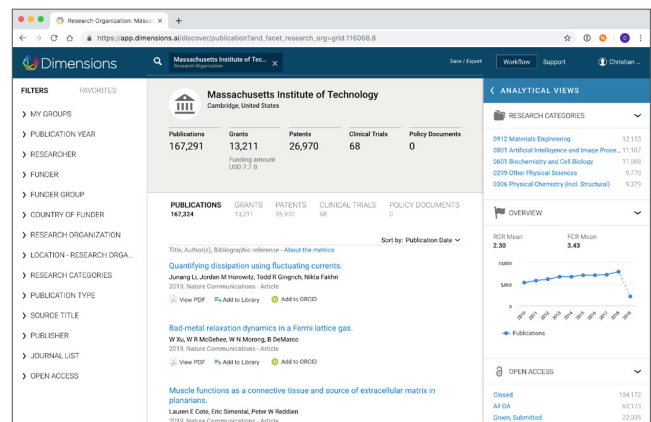
出版物において（研究助成金や特許でも同様ですが）著者は所属機関名をさまざまに表記します。実際、ほとんどの機関はそれほど名前のバリエーションがないのですが、中には数百通りもあったりします。Dimensionsのようなデータのインフラストラクチャにおいては、ドキュメントを正しい機関に自動的に結びつけられるようなユニークIDが必要です。さらに、何をもって機関とするのか、なぜ収録対象とするのか、どのような機関であるのか、明確なポリシーに基づいて、各機関はそのユニークなIDのリストの中できちんと説明されていなければなりません。そして、地理位置情報や創立年月日など有用なメタデータがなければなりません。最も重要なのは永続的識別子です。

Digital Scienceはすでにこの課題に取り組んでおり、その成果としてオープンな[GRIDデータベース](#)をリリースしています。すでに90,000以上の機関についてデータを整理し、各機関には永続的識別子を付与しています。GRIDは継続的に改良されており、例えばORCIDなど他の多くのシステムで利用されています（[ORCIDのブログ](#)を参照）。

Dimensionsでは、異なるデータタイプの中でも各機関の一貫性が保てるよう、GRIDを活用しています



GRID上のMITのレコード



Dimensions上のMITのプロフィール

GRIDを より良いものに するために

私たちがより多くのデータを集めるにしたがって、GRIDデータベースは継続的に改良されています。Digital Scienceは研究コミュニティを支援するため、オープンなパブリック ドメインのデータセットとして今後もGRIDを提供していきます。GRIDはまだ不完全で、今後もずっとそうでしょう。研究機関は合併したり、名称変更したり、常に変わり続けていますし、新しい機関も設立され、あなたが考える以上に流動的なのです。GRIDの詳細については www.grid.ac をご覧ください。また、データ改善のためのサジェスションは、Dimensionsからサポートリクエストをお送りください。



人物同定のチャレンジ：出版物から助成金、特許、臨床試験まで

出版物の著者の同定を自動的に行うのは常に困難がつきまといました。研究コミュニティの中でORCIDを採用する機関が増えてその普及が進んでも、Symplectic Elementsのように大学や助成機関が出版物と研究者、助成金情報のリンクを管理するのを助けるようなソリューションはまだ必要です。Dimensionsチームにとっては、自動で人物同定を行うのはとても難しいのです。Dimensionsでは、助成金、出版物、臨床試験、特許という4種類のソースデータすべてに研究者を結びつけなければならないからです。結果的に、ソースデータに含まれるメタデータだけではなく、ORCIDの公開データを利用して現時点で最善の結果を得るように自動化された研究者同定のプロセス開発に、多くのリソースを費やすことになりました。

技術的な観点から言えば、Dimensionsにおける同定プロセスは、再現率 (recall) よりも適合率 (precision) に重点をおいたアプローチをとっています。データエラーは信頼を損ねて混乱を招きますし、出版物やその他のドキュメントを間違って研究者に結びつけてしまうよりは、不完全なデータである方がまだよいでしょう。完全性を目指すのであれば、ユーザーの助けを得る方が簡単ですし、データベースへの不信感が募るよりは、ユーザー経験上は弊害が少ないと考えます。

引用と謝辞による研究文脈の追加

異なるタイプのコンテンツどうしの引用情報を抽出してリンクすることがDimensionsの要です。分離されたデータの壁を取り払って、ひとつの研究の文脈について優れた洞察をユーザーに提供することが、私たちの目的だからです。そのようにデータをとりまとめることができれば、一つの領域について、関連する研究プロセスも含めてより豊かな背景が見出せるでしょう。そうすれば、以前なら途方もない努力が必要だった新たな知見を得ることができます。

Dimensionsの引用情報は、既存のデータベース（CrossrefやPubMed Central、Open Citation Dataなど）からハーベストしたり、出版社から提供される全文テキストから直接抽出したりしています。これはジャーナル論文の引用だけでなく、謝辞や、学術書からの（あるいは学術書への）引用、会議情報や特許、助成金、臨床試験にいたるまで、すべてを含みます。

出版物だけではない
研究者名の
表記ゆれ

異なるタイプの
コンテンツ間の
参照情報を
抽出

出版物
関連するデータ
参照している文献
他の文献からの引用
関連する助成金
特許からの引用
関連する臨床試験
政策ドキュメントからの引用

特許
参照している文献
関連する助成金
他の特許からの引用
参照している特許
政策ドキュメント
参照している文献

臨床試験
関連する文献
関連する助成金
助成金
助成の成果としての文献
助成の成果としての特許
助成の成果としての臨床試験



13億以上の 相互引用リンク

Dimensionsに収録されたレコードの相互リンクの数は合計で13億、そのうち10億は出版物レコード内のリンクです。より多くの出版社やソースへと収録範囲を拡大するに当たって、抽出されるリンクも増えています。

出版物を超えて： 可能な限り 多くのソースから コンテンツを収録

さまざまな
コンテンツを
リンクする
革新的な
アプローチ

Dimensionsは他のデータベース、例えばGoogle Scholarや Pubmed、Scopus、Web of Science等とどう違うのか？

PubMedやGoogle Scholar、Scopus、Web of Scienceよりも収録対象が広範囲に及ぶDimensionsは、単純に比較できません。出版物とその引用情報という点だけとて、Dimensionsの収録数は他のデータベースよりも25%多く、より包括的です。しかし、Dimensionsが他のツールに勝るのは、出版物に加えて助成金情報や特許、臨床試験情報、政策ドキュメントなどを広く収録し、一貫した相互リンクによって新たな可能性をもたらしているという点です。適切なディスカバリーと研究計画、インパクトのコミュニケーションを可能としているのです。Dimensionsのユーザーインターフェースでは検索結果を一望できると同時に、関連の深い出版物や助成金その他の情報どうしのつながりやインパクトを精査することができます。

Dimensionsを使えば：

- ScopusやWeb of Scienceでも提供される引用情報はもちろん、
- Google Scholarに似たアプローチで、広範囲の情報から全文テキストをベースにした精度と関連度の高い検索を実現できます。
- 出版後ではなく将来の研究を示唆する特許情報で新しいトレンドを把握できます。
- 研究成果だけでなく、そのプロセスを含む広く豊富なコンテンツをもとにして、出版物とその引用情報を超えて、インプットからアウトプット、インパクトと相互の関連性について深い理解が得られます。

DimensionsのパワフルなAPIで、M2M (machine-to-machine)のコミュニケーションを実現します。原則として契約機関向けの有料オプションですが、リクエストをいただければ研究目的の個人利用も可能です。



さまざまなデータベースの引用カウント

Dimensionsについていただくご質問のひとつに、「Google ScholarやScopus、Web of Scienceの引用カウントとどう違うのか？」というのがあります。できるだけシンプルにお答えしたいのですが、それは不可能です。まず、Dimensionsに収録されている参照情報は他のデータベースと単純に比較できません。従来のように論文ベースの引用情報を越えた参照リンクを形成しているからです。出版物の引用カウントだけをとってみても、単純なランクづけは不可能です（Scopusが2006年に提供を開始した直後にも、すでにそのような計量書誌学的な報告がありました）。同じレコードについて、Google ScholarやScopus、Web of Scienceその他のサービスが異なる引用カウントを示すのは、いくつかの理由があります。例えば以下のようなものです：

引用カウントがデータベースによって異なる理由

- 各データベースが収録するソース情報はそれぞれ異なるため、得られる引用情報も違ってきます。
- 各データベースが対象とする収録期間も異なります（例えば1996年以降、など）。
- 各データベースが収録対象とするコンテンツのタイプも異なります。例えば、査読付きジャーナルの情報だけを収録しているものもあれば、ウェブ上に公開された学位論文やプレプリントや電子公開版（同じ論文のプレプリントと出版社バージョンの区別などさらに複雑ですが）のように、出版前のもの、あるいはそもそも出版されていないものからの引用情報まで含むものもあります。
- データベースの更新頻度も、毎日更新されるものから週に一度、あるいはそれ以下など、いろいろです。
- 参照情報をレコードから抽出してそれを引用文献として同定するのはたいへんな作業で、各データベースによってアプローチは異なります。業界として標準的な手順は確立されていませんし、その結果、適切に同定できなかったり、誤って同定されてしまうこともあります。
- 同定アルゴリズムが改良されても、新たなデータソースが加わってデータベース全体が更新されれば、引用カウントも変化します。

Dimensionsのレコードをソースデータに照らして抽出検査してみると、引用カウントが少ない場合もあれば、著しく多い場合もあります。もっと多くの出版社から情報を集めなければならない分野もありますし、より多くの助成金情報が必要なところもあります。同様に、特許情報の収録をもっと充実させなければいけない地域もあります。これについても、研究コミュニティからのフィードバックをいただきながら、収録範囲についての優先順位を決めていきたいと考えています。



Dimensionsと その収録データ： さらなる 品質向上を めざして

より良いデータを 目指して

一例として、PLOS Oneに掲載されたある論文について、各データベースでの被引用数を調べたところ、右表のような結果となりました：

Price MN, Dehal PS, Arkin AP (2010)
FastTree 2 – Approximately Maximum-
Likelihood Trees for Large Alignments.
PLOS ONE 5(3): e9490.
<https://doi.org/10.1371/journal.pone.0009490>

Dimensions:	3,041
Scopus:	2,691
WoS:	2,713
CrossRef:	2,658
Google Scholar:	3,777

先にあげたようなさまざまな理由で、異なるサービスが全く同じ数字を示すことはほとんどありません。結果的に、引用のカウントは単一のシステム内で比較をするために便利なものとして考える人が多いのです。

現在の収録範囲と品質を出発点として

現在収録しているソースとコンテンツをひとつにまとめるには、たいへんな努力とリソースを費やしましたが、まだこれは始まりにすぎません。

- 数ヶ月ごとに新たな助成機関の協力とデータ提供を得て、助成金情報を継続的に追加していきます。
- 2019年は年間を通じて、より多くの出版社、特許庁、臨床試験レジストリ、政策ドキュメント発行機関からデータを追加していきます。
- プレプリントについても一貫して収集する予定です。
- Dimensionsにコンテンツ収録を希望する出版社にもサポートを提供していきます。

最も重要なのは、皆さんからのインプットとフィードバックです。どのように私たちのデータを良くしていくことができるか、皆さんからのご意見をお待ちしています。開発パートナーからすでに多くの課題をいただっていますが、私たちはそれ以上を目指しています。これはまさに共同作業であって、ユーザーの皆さんも、研究コミュニティも、大きなDimensionsチームの一員なのです。



確かな未来の兆候としての 助成金情報

助成金の獲得は、研究者、あるいは研究チームがこれから取り組もうとする研究プロジェクトを記述した詳細なプロセスの結果です。彼らの目的は、研究課題が興味深いもので、実現可能であって価値あるものだということ、そして研究チームに約束された成果を達成する能力があるのだということ、を、助成機関と匿名の審査員たちに納得させることです。このプロセスはたいへん重要なもので、多くの場合、使われるお金は公的資金ですから、十分な説明責任が必要となります。助成金情報は、研究のアイデアの価値を第三者に認めてもらうために、説得力のある形で初めて表明するもので、ソフトウェアに例えればベータ版です。研究サイクルの中で助成金情報は特別な価値をもっています。今後数年間にわたって実施されようとしている研究について知ること、分野のトレンドや動向についての分析が可能となるからです。まさに未来の兆候です。助成機関や研究政策担当者は、助成金の全貌を分析することによって、遡及的な事実確認や誤った決断をチェックするだけでなく、早期の介入や戦略立案にも役立てることが出来ます。

ÜberResearchは、Dimensionsの開発に携わったDigital Scienceの6つのポートフォリオ事業の一つです。研究助成機関が管理する大きな助成金データベースを構築するため、2013年に創業されました。その目的は国および機関レベルで研究システムのインプットの側面を初めて明らかにし、公共のステークホルダーに対して報告するための特注システムを使う大きな助成機関はもちろん、少人数のスタッフと限られたリソースしか持たない小さな機関にまで提供することでした。ÜberResearchの初期の事業は、インプットだけでなく学術的な注目度や事業化、政策立案、インパクトに至るまでカバーする広大なDimensionsの一部として引き継がれています。

助成金情報は取り扱いが難しく、それにはいくつかの理由があります。出版物のように共通したメタデータ構造がなく、DOIのような永続的識別子もまだありません。各国の研究助成の枠組みに大きく依存していて、地理的な差異が大きいのです。ほとんどの研究助成金を競争的資金としている国もあれば、目的を指定しない補助金の制度をとる国もあって、後者は助成金データベースに含まれることはほとんどありません。もちろん、一般補助金と競争的資金が混在する制度をとる国も数多くあります。このため、最近の[レポート](#)でも指摘したように、助成金データは研究関連資金の全容を示すものとしてとらえるべきではなく、政府や国際機関、財団などさまざまなタイプの助成機関が提供する研究プロジェクト資金として考えるべきです。助成金情報を利用するにあたってご質問があれば、ご遠慮なくお知らせください。

助成金データは
将来の予兆を示す
貴重な情報源

ÜberResearchは
1.5兆ドルの
助成金を管理する
データベースを
構築

全体ではなく
個々の特徴を
とらえる



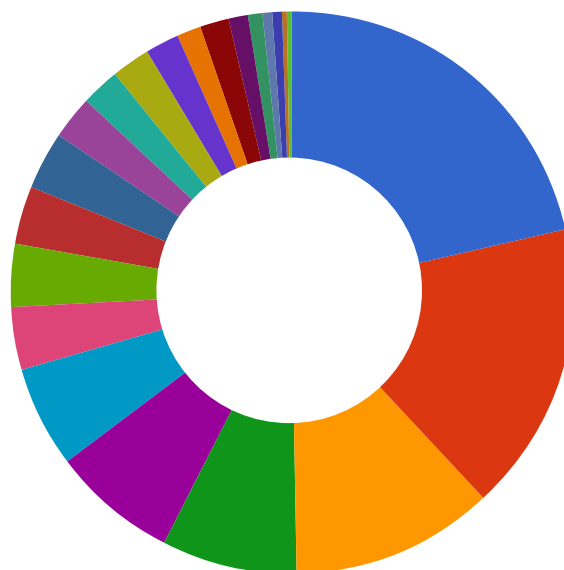
Dimensionsの助成金データ

以下は2019年4月1日現在の情報です。データは毎月更新されていますので、ここに示す情報と実際にDimensionsやそのAPIからの検索結果とは異なることにご注意ください。

助成金レコード	4,600万件
収録助成機関数	340機関以上
助成金総額	1.5兆ドル
1件あたりの平均助成額	40万3,000ドル
2019年以降アクティブな助成金総額	3,410億ドル
助成金獲得機関数 (GRID IDへのリンク)	4,600万機関
助成金獲得研究者数 (研究者IDへのリンク)	6,300万件

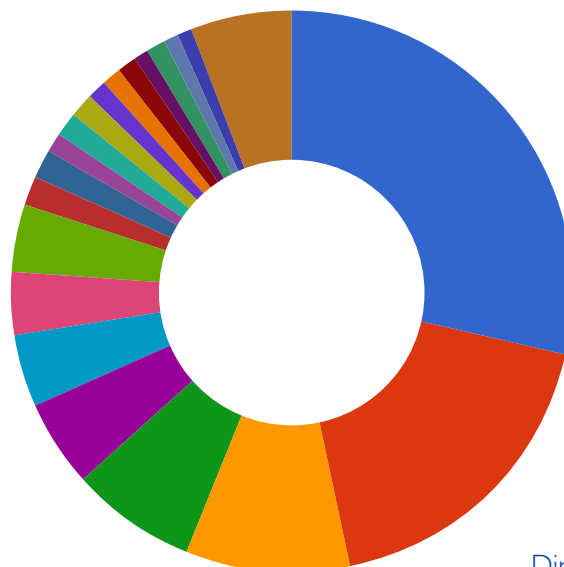
助成金データの分野別分布状況

- Medical and Health Sciences
- Biological Sciences
- Engineering
- Chemical Sciences
- Information and Computing Sciences
- Physical Sciences
- Mathematical Sciences
- Psychology and Cognitive Sciences
- Studies in Human Society
- Earth Sciences
- History and Archaeology
- Environmental Sciences
- Education
- Language, Communication and Culture
- Economics
- Technology
- Commerce, Management, Tourism and Services
- Agricultural and Veterinary Sciences
- Law and Legal Studies
- Philosophy and Religious Studies
- Built Environment and Design
- Studies in Creative Arts and Writing



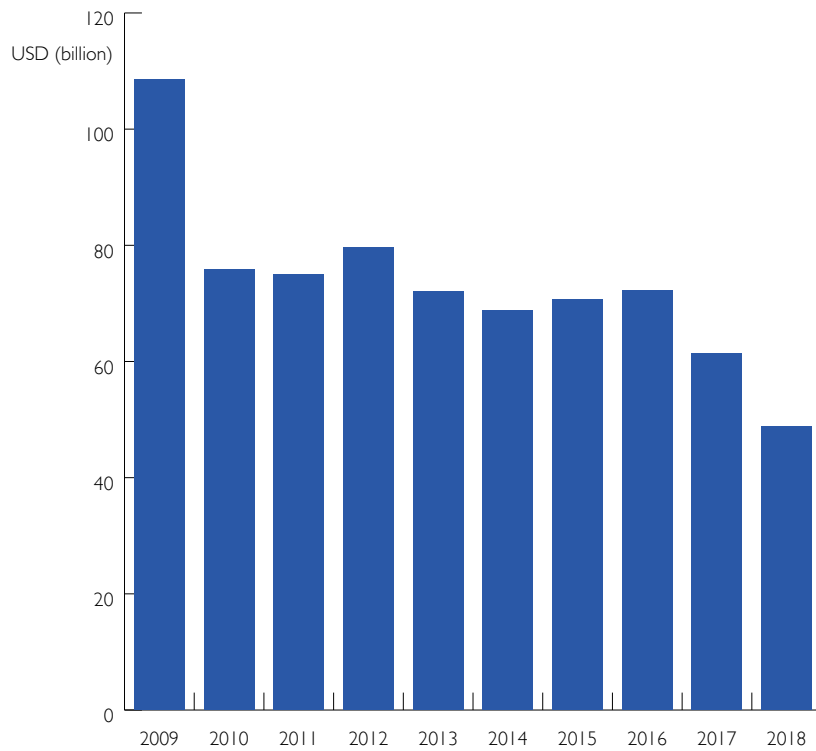
助成金データの地域別分布状況

- United States
- Japan
- Canada
- Germany
- China
- United Kingdom
- Russia
- South Africa
- Switzerland
- Brazil
- Australia
- Poland
- South Korea
- Sweden
- France
- Czechia
- Italy
- Belgium
- Netherlands
- Norway
- Other

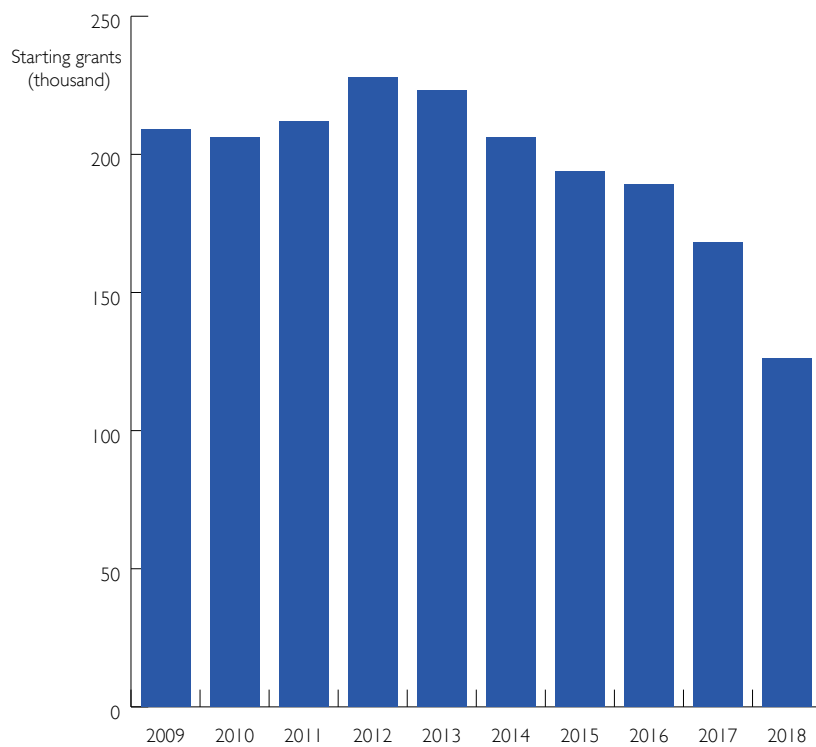




助成金総額（助成開始年による）



助成金件数（助成開始年による）





出版物と引用情報

寡占市場が
もたらした
イノベーションの
欠落

Dimensions :
模倣ではない
新たなアプローチ

包括性を目指して：
ベンダーではなく
ユーザー本位の
収録方針

Dimensionsの出版物データと引用情報： データベースは「審査基準」ではありません

パワフルな書誌引用データベースとしてのDimensionsは、より一層のメタデータへのアクセスと利用を、研究者と研究機関に提供することを目的として開発されました。これはDigital Scienceの悲願でした。データベースの寡占市場は多くの点で、研究者を支援するために必要なイノベーションに遅れをとってきました。研究者や研究支援者、評価者のニーズに応える活発で研究主導型のディスカッションではなく、技術的な面でも実用面でも、また研究政策から生じる特定の状況においても、旧来の制約にしばられた狭量なアプローチに終始していました。

単純に既存のアプローチを繰り返して第3の（あなたが何をどうカウントするかによっては第4、第5の）書誌索引データベースを開いても、市場の期待に応えられないのは明らかでした。このため、私たちは根本的に異なる以下の2つのアプローチをとることに決めたのです。

- Dimensionsは研究に関連するすべての成果物を統合すべきであって、本質的に収録対象を選択するべきではありません（ハゲタカジャーナルなどはもちろんその限りではありません）。
- 出版物だけでなく、助成金や特許など他のソースも同様に扱い、一貫してリンクすべきです。

Dimensionsというデータベースをできる限り包括的なものにするというのが私たちの主な関心事です。技術の進歩によってユーザーからの期待は高まっていると私たちは強く感じています。提供事業者の都合によって収録対象を制限することなど、もう誰も望んでいないと思います。実際、このレポートの執筆中にも、ネットワーク中立性が現実社会で大きな問題となっています。私たちが収録・提供するコンテンツについても、意識的に中立でなければならないと考えます。つまり、何がデータベースに収録すべき「価値ある」アウトプット（例えばジャーナルなど）かという判断は、私たちではなく、研究コミュニティやそれを使う個人に任せるべきなのです。むしろ私たちの仕事は、コンテンツを検索して最も関連性の高い結果を最も効率的に引き出すための最良のツールを提供することであると信じています。

品質保証のフィルタリングとしてのホワイトリスト/ブラックリスト

ユーザーがその用途に応じて判断し、適切なコンテンツを選択できるよう、Dimensionsには検索結果の絞り込みに使えるサブセットが設定されています。あらかじめ定義され、キュレーションされたリストを標準のフィルターとすることで、ホワイトリストとしてもブラックリストとしても使うことができます。まず私たちが採用したのは、研究コミュニティの中ですで受け入れられ、自由に使うことができるリストですが、今後さらにコミュニティからのご意見をいただければ幸いです。



当初よりDimensionsに採用したのは、以下のようなリストです。

- DOAJリスト : Directory of Open Access Journals (DOAJ) は、コミュニティによって維持されているオンラインディレクトリで、高品質な査読付きのオープンアクセスジャーナルを対象としています。10,000誌以上に上るDOAJの収録ジャーナルは、科学技術、医学、社会科学や人文学まで広くカバーしています。
- ERAリスト : ERA 2015年版ジャーナルリストは、豪州研究会議 (Australian Research Council, ARC) が国家保健医療研究会議 (National Health and Medical Research Council, NHMRC) および広く研究コミュニティからの協力を得て作成したものです。オーストラリアの国家的研究評価の枠組みであるExcellence in Research for Australia (ERA)での利用を目的として、研究機関がERA 2015への評価資料提出に含めることができるジャーナルを定義しています。ERA 2018年版についてもリリースされ次第、採用する予定です。
- ノルウェー・レジスター : ノルウェー・レジスター (Norwegian Register for Scientific Journals, Series and Publishers) は、ノルウェー研究データセンター (Norwegian Centre for Research Data, NSD) と国家学術出版委員会 (National Board of Scholarly Publishing, NPU) が共同で管理しています。ノルウェーの研究資金配分の重み付けで使われる約30,000誌の学術出版物を定めています。
- PubMedリスト : PubMedは主にMEDLINEをソースとする生命科学・生物科学系の出版物情報を収録するデータベースです。米国の国立衛生研究所 (National Institutes of Health, NIH) が管轄する国立医学図書館 (National Library of Medicine, NLM) によって運用されています。DimensionsでPubMedフィルターを使うと、PubMedで規定されているID (PMID) の付与された出版物に絞り込むことができます。

既存の
ジャーナル・リスト
を使って
フィルタリング

これらは出発点に過ぎず、特定の用途にしか適しません。もっと汎用的な、あるいは特定の国や機関が用意したリストや、その他のリストが有用だと思われるさまざまなケースについてフィードバックをお寄せいただければ、開発の参考とさせていただきます。

他にも有用な
リストがあれば
お知らせください

Dimensionsの出版情報と引用データの統合

Dimensionsの出版情報と引用データは複雑なプロセスを経て統合されています。「なぜこんな風にデータが見えるのか」理解されたい方のために、以下のとおり2つのステップに分けて重要なポイントをご説明します。

ステップ 1: 屋台骨を作る

まず大掛かりなメタデータの骨格を構築し、継続してアップデートしています。言ってみればデータの屋台骨ですが、これは、PubMedやPubMed Central、Crossrefなど、オープンあるいは最小限のライセンス許諾で利用可能なソースから得られるデータを統合したものです。



1億件の 出版情報を収録

その結果、約1億件のレコードを含むユニークに同定された大きなインデックスが得られました。そのかなり多くの部分はCrossrefから得たDOIが付与された出版情報で、Crossrefメンバーである12,442もの出版社が提供しているものです。これだけでもかなりしっかりとしたメタデータの骨格ができますが、メタデータの完全性という意味では、例えば著者の所属情報のようにまだ欠落しているデータがあるのです。

ステップ 2: データを強化する

ステップ1で得られたメタデータは、入手できる限りの全文テキスト情報によって強化され、コンテンツの発見可能性を飛躍的に高めます。

100以上の 出版社が 6,900万件の 全文テキストを 提供

このステップでは、全文テキストから参考文献情報を抽出したり、謝辞 (acknowledgements) の部分をマイニングして、助成金プロジェクトや助成機関、臨床試験などにリンクさせます。すでに6,900万件の全文テキストについてこの処理を完了していますが、オープンアクセスのものを除けば、大部分はこの作業を目的としてDigital Scienceに提供されたものです。こうしたレコードは、世界最大規模のSTM出版社を含む100社以上から提供されました。Dimensionsを検索していただければ、どの出版社が収録されているかは一目瞭然です。

このデータ強化のステップで肝心なのは、Dimensionsが全文テキストをソースとしているという点です。つまり、ユーザーは論文タイトルや抄録に現れない単語についても検索ができるのです。絞り込み機能を一緒に使うことによって、探し求めている研究情報が見つかる可能性が高まるというわけです。

より多くの コンテンツを 追加

より多くの出版社の協力を得て、さらに多くの出版情報が追加され、発見可能性が高まります。当初の12ヶ月はDimensionsのリリースに向けて、主に大規模から中規模の出版社に重点をおきました。Dimensionsへの収録を希望される出版社は、[こちらのフォーム](#)からご連絡ください。

アカデミックな注目を超えて : Dimensionsのオルトメトリクス

オルトメトリクス : 即時的な インパクト

Digital Scienceはオルトメトリクスの黎明期からこれを支持し、私たちのAltmetricはその発展に重要な役割を果たしてきました。実際、カラフルなAltmetricドーナツやスコア表示、政策ドキュメントや大学シラバスなどのユニークなソース、とても人気のあるAltmetric Top 100など、数々のイノベーションでAltmetricは常に先駆者となりました。

DimensionsはハイレベルなAltmetricデータを統合し、レコード詳細とともに表示します。これにより、アカデミックな注目（引用）、イノベーションに与えた影響（特許）、臨床的なインパクト（治験）のほか、公共政策からの注目などが、Altmetricが収集するSNSをはじめその他多くの指標から明らかとなるのです。

多くの国々で、研究インパクトの実証が必要とされていて、このようにインパクトへの道のりを示すデータをまとめ上げる努力に繋がりました。AltmetricのデータをDimensionsに標準装備することで、研究コミュニティがインパクトを定量的に理解する一助となれば幸いです。



オープンアクセス、オープンな引用データ、そしてDimensions

Digital Scienceはオープンアクセスを熱心に支援しており、コミュニティがそれを促進するにあたってDimensionsが役立つと信じています。包括的なオープンアクセス出版物の情報を提供できるよう、Unpaywallのオープンアクセス情報を統合し、DOAJから得た情報に基づいてOAジャーナルのリストを管理しています。Dimensions無料版でもほとんどのOA論文にワンクリックでアクセスできます。Dimensions上でReadCubeが連携して全文テキストを開くので、ユーザーは迅速にコンテンツにたどり着けます。

引用データを含むメタデータの公開によって、革新的なソリューションやツールの開発が促進された好事例がDimensionsです。Dimensions自体もそれを念頭に開発されました。つまり、一貫してリンクされた高品質のメタデータを提供してアクセスを保証するだけでなく、創造性をかきたてるということです。これだけデータがあれば、研究を支援するイノベーションが可能です。

私たちはたいへん重要な取り組みであるInitiative for Open Citations (I4OC) に賛同しています。Dimensionsは、引用データが自由に使えるようになればどれだけのことができるのかという実例です。Dimensionsの開発にあたって、Digital Scienceは高品質なディスカバリー体験をユーザーに提供できるよう、引用データの構造化に相当な投資をしなければなりません。I4OCや関連するその他のイニシアチブが、継続してこの障壁に取り組んでいくように願っています。そうすればコミュニティは研究を迅速に進めたいユーザーにとって価値ある機能を開発することにもっと注力できるでしょう。

「Digital Scienceは出版社ではないので、引用データについてI4OCに貢献するにはふさわしくない」とよく言われますが、確かに出版社こそがもっと貢献すべきでしょう。Dimensionsだけでなく、AltmetricやFigshareもI4OCに貢献しています。

Dimensionsにおけるオープンアクセス

Dimensionsとオープンな引用情報

Dimensionsの出版・引用データ

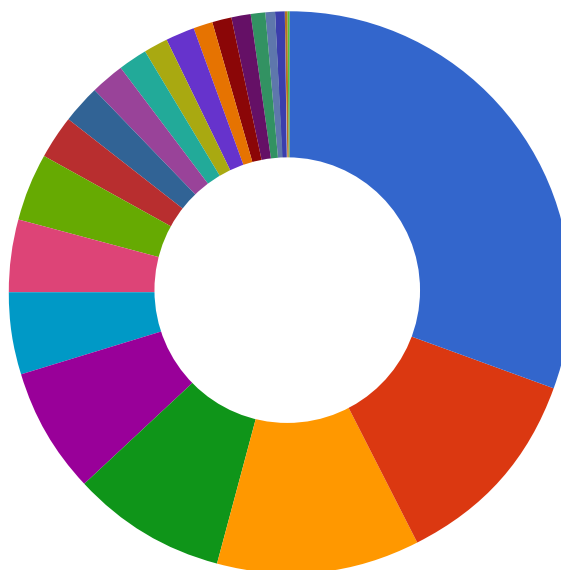
以下は2019年4月1日現在の情報です。データは毎日更新されていますので、ここに示す情報と実際にDimensionsやそのAPIからの検索結果とは異なることにご注意ください。

出版情報	1億件
収録ソース数（ジャーナル、学術書、プレプリント、会議録など）	50,000タイトル以上
著者の所属機関数（GRID IDへのリンク）	1億5,800万機関
著者数（研究者IDへのリンク）	2億900万人
参考文献数	11億件
助成金情報へのリンク数	1,100万件
助成機関へのリンク数	1,700万件
臨床試験情報へのリンク数	891,000件



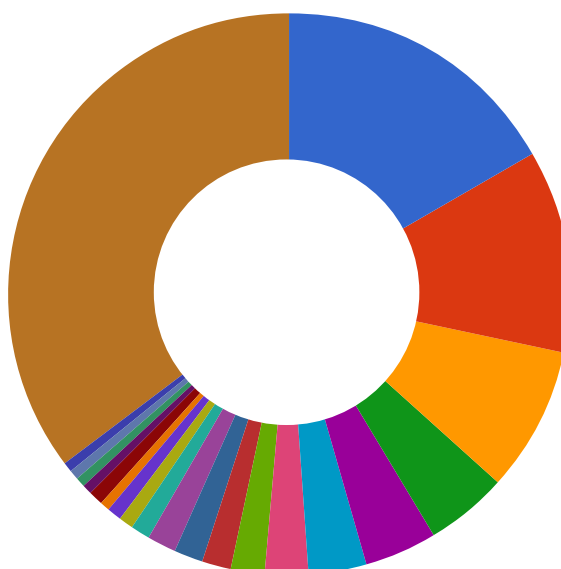
出版物データの分野別分布状況

- Medical and Health Sciences
- Biological Sciences
- Engineering
- Chemical Sciences
- Information and Computing Sciences
- Physical Sciences
- Mathematical Sciences
- Psychology and Cognitive Sciences
- Studies in Human Society
- Earth Sciences
- Technology
- Economics
- Environmental Sciences
- History and Archaeology
- Language, Communication and Culture
- Education
- Commerce, Management, Tourism and Services
- Agricultural and Veterinary Sciences
- Law and Legal Studies
- Philosophy and Religious Studies
- Studies in Creative Arts and Writing
- Built Environment and Designing



出版物データの出版社別分布状況

- Elsevier
- Springer Nature
- Wiley
- Taylor & Francis
- IEEE
- Oxford University Press
- SAGE Publications
- Wolters Kluwer
- JSTOR
- Cambridge University Press
- American Chemical Society
- De Gruyter
- BMJ
- IOP Publishing
- AIP Publishing
- Thieme
- Royal Society of Chemistry
- American Medical Association
- American Physical Society
- SPIE
- Other





臨床試験情報： 臨床応用に向けた研究成果

正確を期すために、「臨床試験」についてWHOの定義を引用します：臨床試験とは健康上の効果を測定するために、患者ないし患者グループに対して一定の治療行為を実施する研究です。（中略）治療行為には以下のようなものが含まれます：投薬、細胞その他の生物学的製剤、外科手術、放射線治療、装置、行動療法、ケアプロセスの変更、予防的治療など。

さまざまな
レジストリから
統合された
臨床試験情報

Dimensionsでは複数の臨床試験登録レジストリからの情報を一度に検索できるようにしています。2019年4月現在収録しているのは以下の10レジストリですが、将来さらに追加の予定です。

レジストリ名	国・地域
ClinicalTrials.gov	米国
UMIN-CTR	日本
EU-CTR	EU諸国
ISRCTN	国際
CTRI	インド
ANZCTR	オーストラリア・ニュージーランド
CHICTR	中国
GCTR	ドイツ
NTR	オランダ
CRIS	韓国

これらの情報はDimensionsの一貫したデータモデルに統合され、カテゴリや研究機関名、出版年などすべてのコンテンツに共通する絞り込みフィルターにマッピングされます。

Dimensionsの臨床試験情報

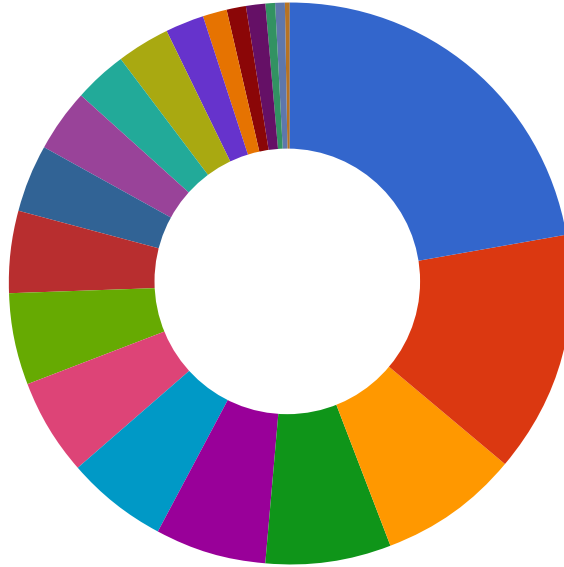
以下は2019年4月1日現在の情報です。データは毎日更新されていますので、ここに示す情報と実際にDimensionsやそのAPIからの検索結果とは異なることにご注意ください。

臨床試験情報	455,000件
収録レジストリ数	10
スポンサー・協力機関数 (GRID IDへのリンク)	130万件
出版情報へのリンク数	441,000件
助成金情報へのリンク数	22,000件
助成機関へのリンク数	571,000件



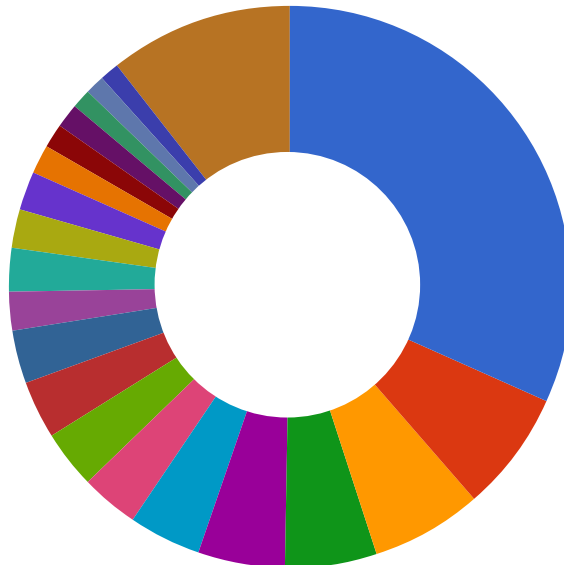
臨床試験データの分野別分布状況 (英国HRCS分類による)

- Cancer
- Cardiovascular
- Metabolic and Endocrine
- Infection
- Mental Health
- Oral and Gastrointestinal
- Musculoskeletal
- Neurological
- Respiratory
- Reproductive Health and Childbirth
- Stroke
- Inflammatory and Immune System
- Renal and Urogenital
- Eye
- Skin
- Generic Health Relevance
- Injuries and Accidents
- Blood
- Congenital Disorders
- Ear
- Other



臨床試験データの分野別分布状況

- United States
- Japan
- United Kingdom
- China
- Germany
- France
- Canada
- India
- Netherlands
- Australia
- Switzerland
- South Korea
- Italy
- Spain
- Belgium
- Denmark
- Brazil
- Israel
- Sweden
- Taiwan
- Other





特許情報： 実用・商用利用に至った 研究成果

Dimensionsのリリース時には主要特許庁の収録から着手しましたが、すでに範囲を拡大して2019年中にはより多くの国と地域がカバーされるようになります。Dimensionsの特許情報の収録にあたっては、研究助成金がどのように研究成果の応用可能性と商業保護に影響を与えているのか、という下流分析を可能にすることを念頭においています。

研究活動から
商用利用への
技術移転を示す
特許データ

特許情報ソース	国・地域
United States Patent and Trademark Office (USPTO)	米国
European Patent Office (EPO)	欧州
World Intellectual Property Organisation	国際
German Patent and Trademark Office (DPMA)	ドイツ
Canadian Intellectual Property Office (CIPO)	カナダ
Intellectual Property India (IPI)	インド
Intellectual Property Office (IPO)	英国
National Industrial Property Institute	フランス
Intellectual Property Department (IPD)	香港
Russian Patent Office	ロシア

Dimensionsの特許情報

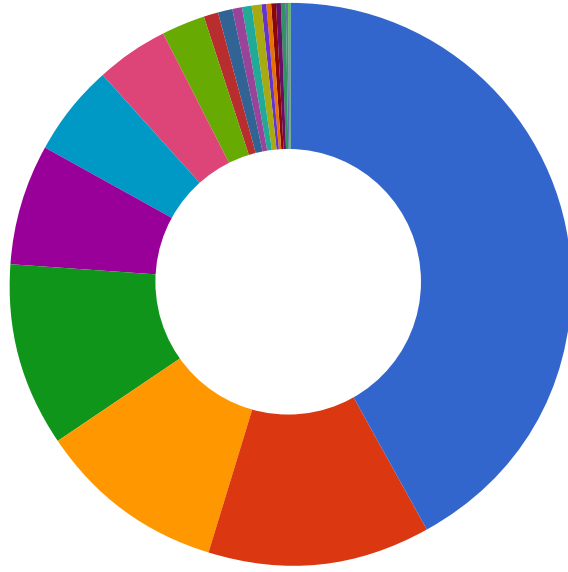
以下は2019年4月1日現在の情報です。データは週に一度更新されていますので、ここに示す情報と実際にDimensionsやそのAPIからの検索結果とは異なることにご注意ください。

特許情報	3,800万件
収録特許情報ソース	10
発明者・出願人の所属機関数 (GRID IDへのリンク)	3,700万件
被引用特許件数	2億2,700万件
出版情報へのリンク数	1,000万件
助成金情報へのリンク数	165,000件
助成機関へのリンク数	221,000件



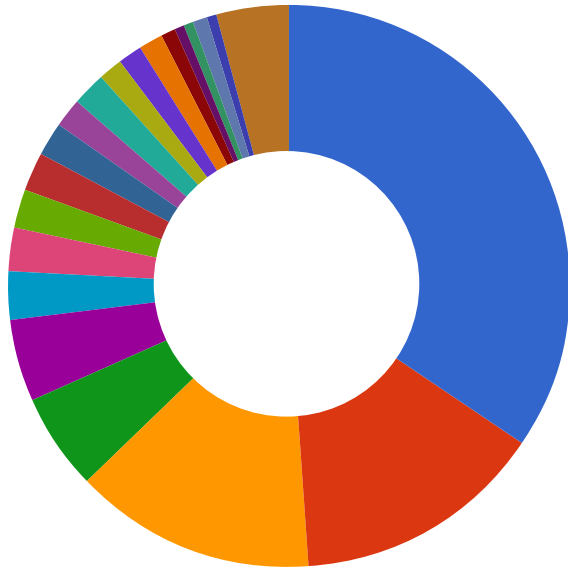
特許データの分野別分布状況

- Engineering
- Information and Computing Sciences
- Medical and Health Sciences
- Chemical Sciences
- Biological Sciences
- Technology
- Physical Sciences
- Psychology and Cognitive Sciences
- Studies in Human Society
- Mathematical Sciences
- Language, Communication and Culture
- Earth Sciences
- History and Archaeology
- Agricultural and Veterinary Sciences
- Studies in Creative Arts and Writing
- Economics
- Environmental Sciences
- Commerce, Management, Tourism and Services
- Built Environment and Design
- Law and Legal Studies
- Philosophy and Religious Studies
- Education



特許データの地域別分布状況（出願人による）

- United States
- Japan
- Germany
- France
- United Kingdom
- South Korea
- Switzerland
- Russia
- Netherlands
- China
- Canada
- Australia
- Italy
- Sweden
- Taiwan
- Finland
- India
- Austria
- Belgium
- Israel
- Other





政策ドキュメント： 政策やガイドラインに 影響を与えた研究

Dimensionsの政策ドキュメント情報は、Digital Scienceのポートフォリオ企業のひとつであるAltmetricが提供しています。ガイドラインや政策とその実践に変化あるいは影響をもたらしたような政策ソースが含まれます。例えば政府が発行するガイドラインやレポート、ホワイトペーパー、独立政策機関による刊行物、特定のトピックや組織や国際開発機関についての諮問機関、などがあります。気候変動や健康、運輸、経済など、できるだけ広範囲の政策ソースを世界中の機関から収録したいと考えています。また、可能な限り全文テキストをもとに索引付けして、分類と参考文献の抽出を行います。

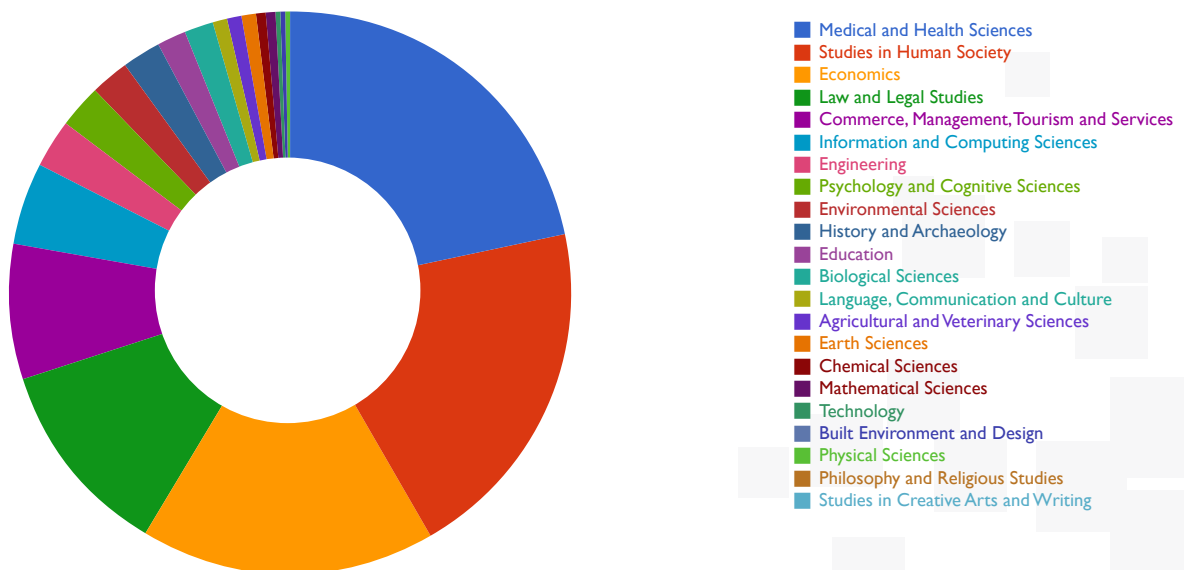
70以上の機関から
政策関連文書を
収集

Dimensionsの政策ドキュメント情報

以下は2019年4月1日現在の情報です。データは毎日更新されていますので、ここに示す情報と実際にDimensionsやそのAPIからの検索結果とは異なることにご注意ください。

政策ドキュメント	421,000件
収録機関数	72機関
出版情報へのリンク数	150万件

政策ドキュメントデータの分野別分布状況





Thank You

Dimensionsにご関心をお寄せいただきありがとうございます。研究コミュニティと皆さんからのご協力をいただき、より良いツールとデータの開発に今後も努めてまいります。

本レポートの執筆にあたってはできる限り正確な記述を心がけましたが、記載している情報については執筆時点のものであり、ご契約にあたっての根拠とはなりませんのでご注意ください。また、第三者に関するいかなる記述についても、DimensionsもしくはDigital Scienceによる、またはそれらに対する賛同や後援、協力を示すものではありません。

Part of **DIGITAL**science



www.dimensions.ai

digital-science.com