

Open Science and the PID Graph

J-Stage Seminar, Japan

October 27, 2020



オープンサイエンスと PID Graph

J-Stage Seminar, Japan

October 27, 2020

Open Science



*“Open Science for its own sake has never been the goal. While a focus on Open Science as a mechanism must be emphasised in any transition, **Open Science must ultimately be embedded as part of a larger more systemic effort to foster all practices and processes that enable the creation, contribution, discovery and reuse of research knowledge more reliably, effectively and equitably.** Research cannot be ‘excellent’ without such attributes at its core.”*

Source: <http://doi.org/10.2777/00139>

Progress on Open Science: Towards a Shared Research Knowledge System

Final Report of the Open Science Policy Platform

Eva Mendez (Chair, OSPP Mandate 2), Rebecca Lawrence (Editor and Coordinator),
Catriona J. MacCallum (Writing Chair), Eva Moar (Rapporteur), together with all the
members of the Open Science Policy Platform

オープンサイエンス



“オープンサイエンスは、それ自体を目的としたものではない。どのような移行においても、メカニズムとしてのオープンサイエンスに焦点を当てることは強調されなければならないが、オープンサイエンスは最終的には、研究知識の創造、貢献、発見、再利用をより確実かつ効果的かつ公平に可能にするすべての実践とプロセスを促進するための、より大きな体系的な取り組みの一部として組み込まれていなければならない。このような特性を核としたものでなければ、研究は「優れた」ものにはなりえない”

Progress on Open Science: Towards a Shared Research Knowledge System

Final Report of the Open Science Policy Platform

Eva Mendez (Chair, OSPP Mandate 2), Rebecca Lawrence (Editor and Coordinator), Catriona J. MacCallum (Writing Chair), Eva Moar (Rapporteur), together with all the members of the Open Science Policy Platform

出典: <http://doi.org/10.2777/00139>

Collaboration & infrastructure



Japan Link Center (JaLC), DataCite and Crossref are Registration Agencies with the DOI Foundation.

JaLC is a member of DataCite and Crossref to support workflows for their members.

DataCite have had a longstanding relationship with JaLC and continue to work collaboratively to support our respective members.

Together as key stakeholders in the scholarly infrastructure community we support interoperability across PID infrastructure.



連携とインフラ



Japan Link Center (JaLC)、DataCite、CrossrefはDOI財団の登録機関である。

JaLCはDataCiteとCrossrefの会員向けワークフローをサポートしている。

DataCiteはJaLCと長年の関係を築き、それぞれのメンバーをサポートするために協力関係を維持している。

学術インフラコミュニティの主要なステークホルダーとして、我々はPIDインフラ全体での相互運用をサポートしている。



Community



It is critical that the libraries, research institutions, and data centers that house repositories have a voice in decisions and receive the optimal services and support.

DataCite is a member organization providing **DOI registration services** and **research infrastructure** for the research community. Our community is the core to everything that we do.



コミュニティ



リポジトリを保持する図書館、研究機関、データセンターが意思決定に発言権を持ち、最適なサービスとサポートを受けることが重要である。

DataCiteは、研究コミュニティにDOI登録サービスと研究インフラを提供する会員組織である。コミュニティが活動の核となっている。





33 consortia

2,193 repositories

238 members

600+ consortium organizations

20M+ DOIs

43 countries



33 コンソーシアム

2,193 リポジトリ

238 メンバー

600+ コンソーシアム 機関

20M+ DOI

43 ヶ国

DOI registration



DataCite's core activity is providing tools for DOI and metadata registration. DOIs are registered for content that is hosted in repositories (or in some cases journals or other containers for hosting content). Thus there is an important and natural relationship between the content (whether it is a dataset, text document, sample, etc.) and the repository.



DOI 登録



DataCiteの中心的な活動は、DOIとメタデータ登録のためのツールを提供すること。DOIは、リポジトリ（あるいはジャーナルやその他のコンテンツを収載するためのコンテナ）に登録されているコンテンツに対して登録される。このように、コンテンツ（データセット、テキスト文書、サンプルなど）とリポジトリの間には、重要かつ自然な関係がある。



PIDs and metadata



It is important that the focus is not exclusively “PIDs” - metadata and services that the PIDs are part of is more important than just the PIDs.

We should coordinate with the broader PID ecosystem - ensuring relations in the metadata



PIDとメタデータ



「永続的識別子 (PID)」だけに焦点を当てないことが重要だ。ーPIDがその一部となっているメタデータやサービスの方が、PIDそのものよりも重要となる。

我々は、メタデータ内の関係を保証するための、より広範なPIDエコシステムと連携すべきである。



PIDs and FAIR



.....PIDs (and the associated metadata) are an essential component for the implementation of the FAIR principles



Findable. Standardized metadata connected with PIDs make research data findable.



Accessible. Resolvable worldwide with every internet browser. URL associated with the DOI can be updated, DOI remains unchanged.



Interoperable. Standard vocabularies and links to other PIDs e.g. software DOIs, instrument DOIs, ORCIDs in the metadata record of a PID.



Reusable. Cite your research sources with confidence, and receive proper credit when your work is reused. Rich and up-to-date metadata generates trust. Linking to other PIDs.

PIDs and FAIR



..... PID（および関連するメタデータ）はFAIR原則の実施に不可欠な要素。



Findable (見つけられる)。PIDと連携した標準化されたメタデータは研究データを検索可能にする。



Accessible (アクセスできる)。すべてのインターネットブラウザで世界中で解決可能。DOIに関連付けられたURLは更新することができるが、DOIは変更されない。



Interoperable (相互運用できる)。標準的なボキャブラリーと他のPIDへのリンク（PIDのメタデータレコード内のソフトウェアDOI、インストゥルメントDOI、ORCIDなど）。



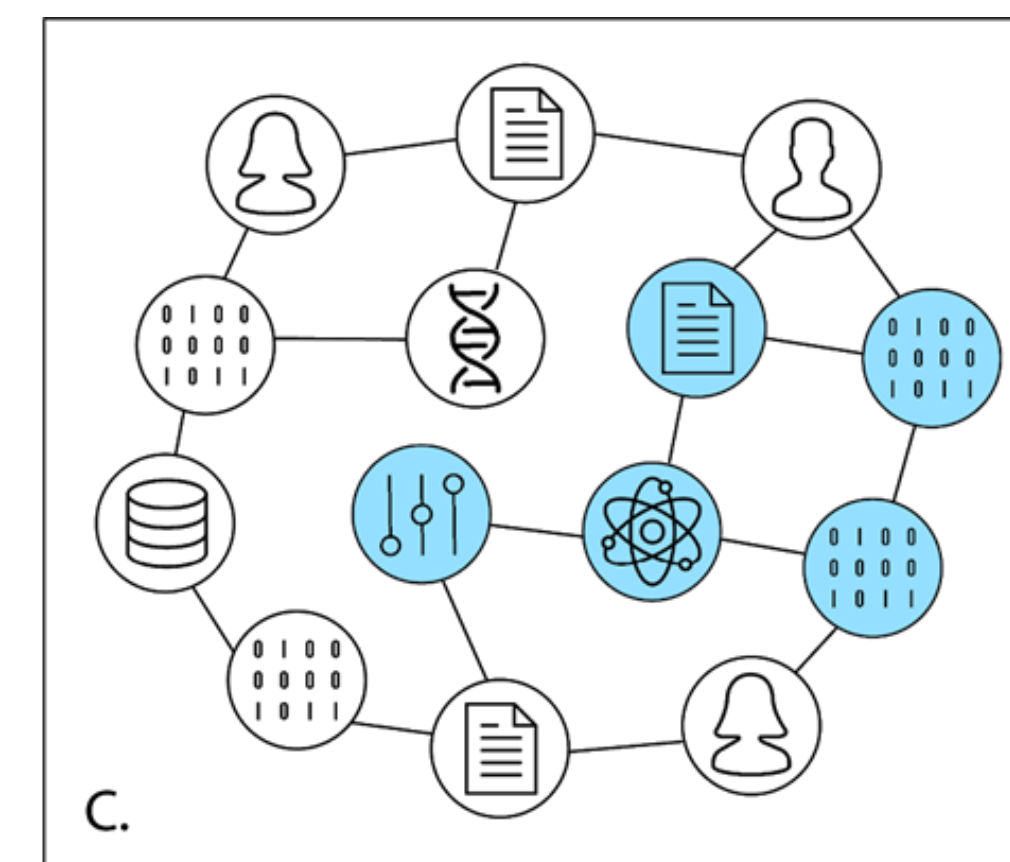
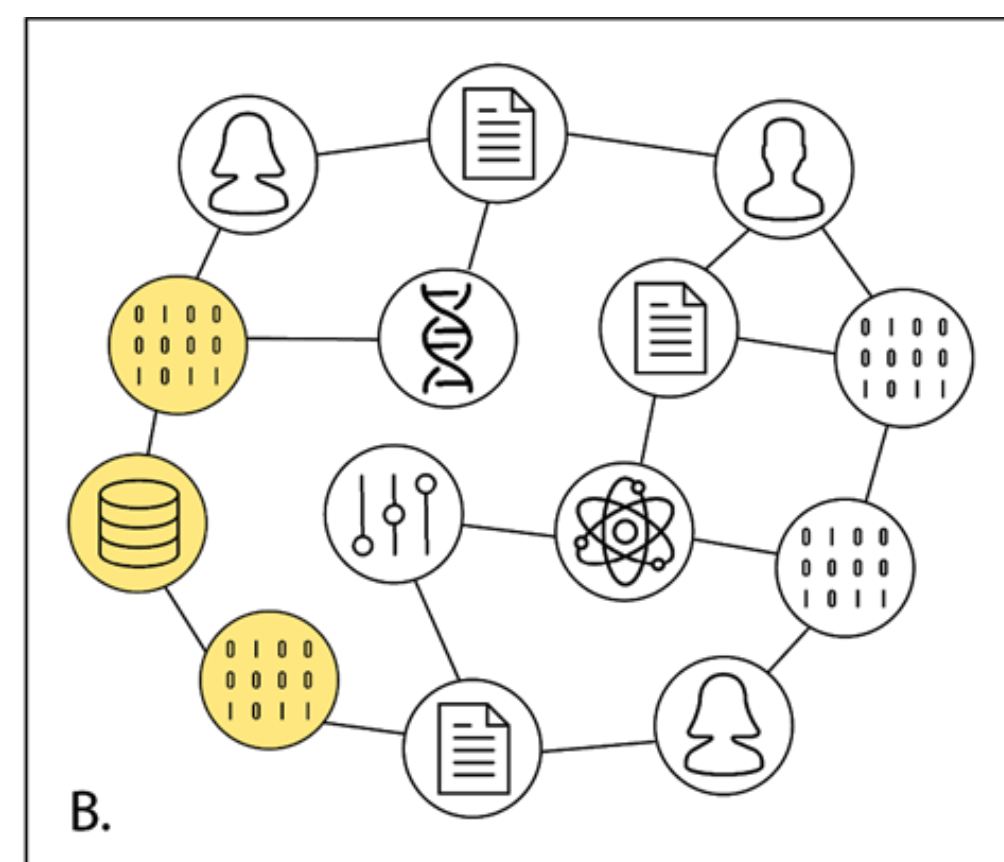
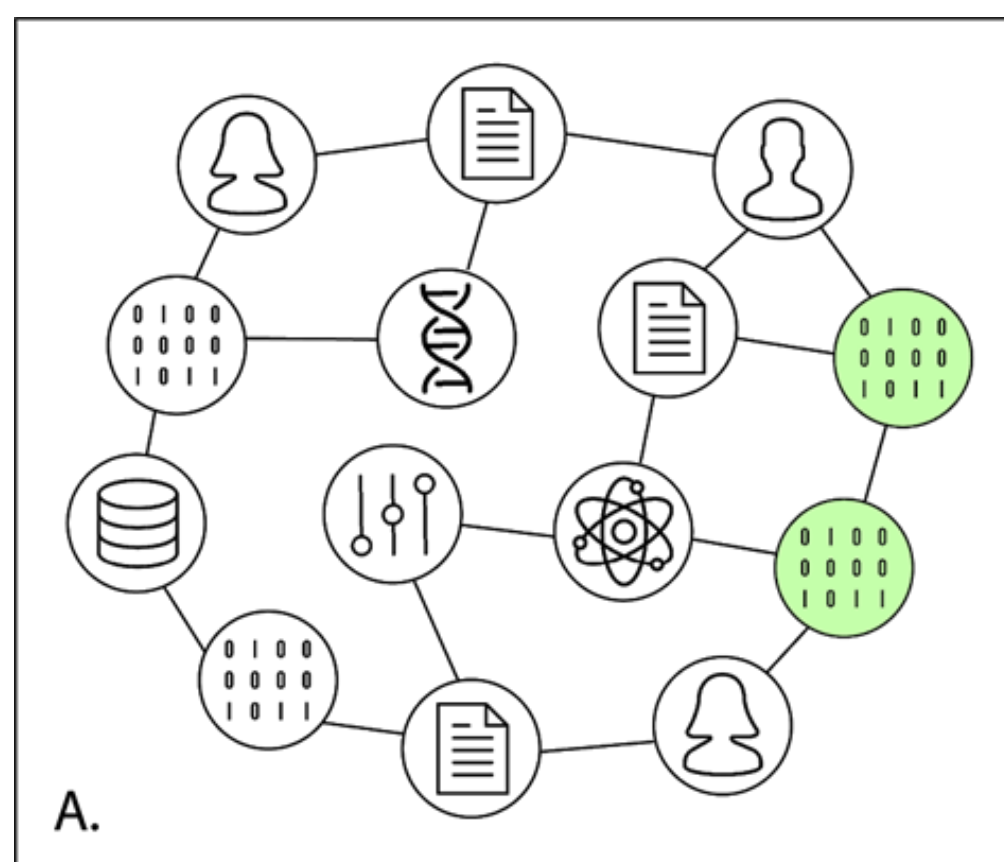
Reusable (再利用できる)。研究ソースを自信を持って引用し、自身の作品が再利用された場合には適切なクレジットを得ることができる。豊富かつ最新のメタデータは信頼を生み出す。他のPIDとのリンク。

The PID Graph



DataCite continue to work on connecting existing persistent identifiers to each other in standardized ways, e.g. to the outputs associated with a particular researcher, repository, institution or funder, for discovery and impact assessment.

To address these use cases we need a more complex model to describe the resources that are identified by PIDs, and the connections between them: a graph. In graph theory, the resources identified by PIDs correspond to the nodes in this graph, and the connections between PIDs correspond to the edges.

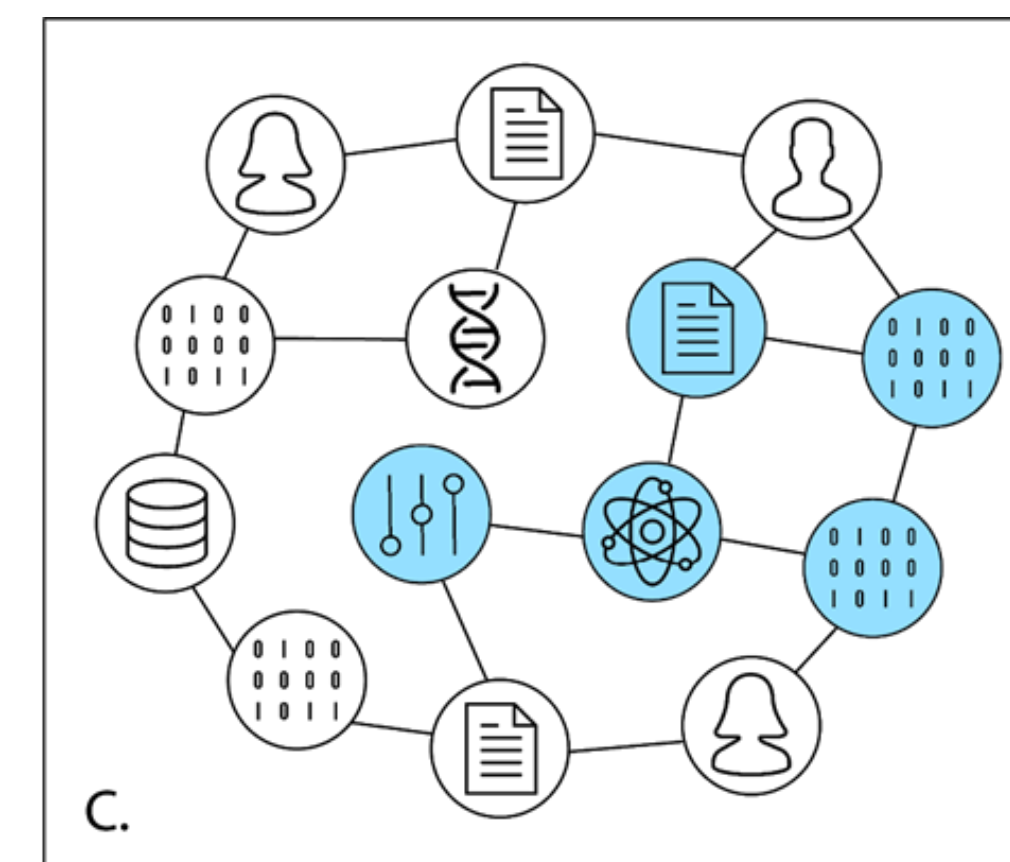
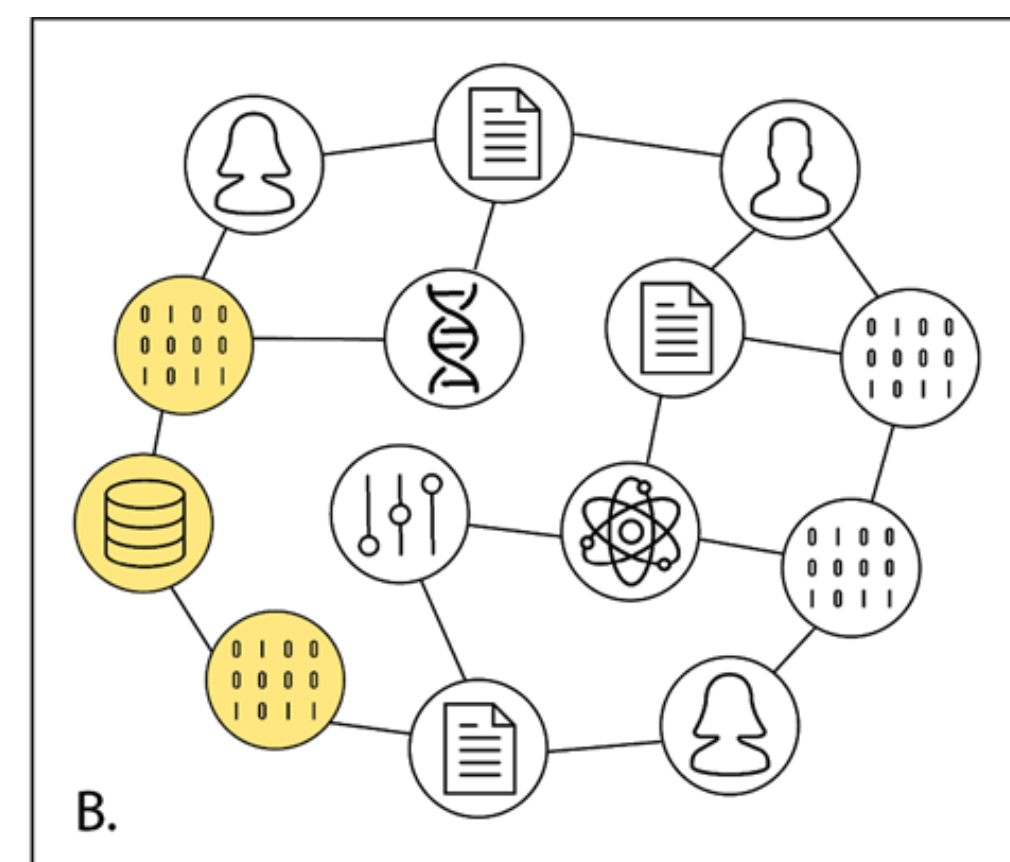
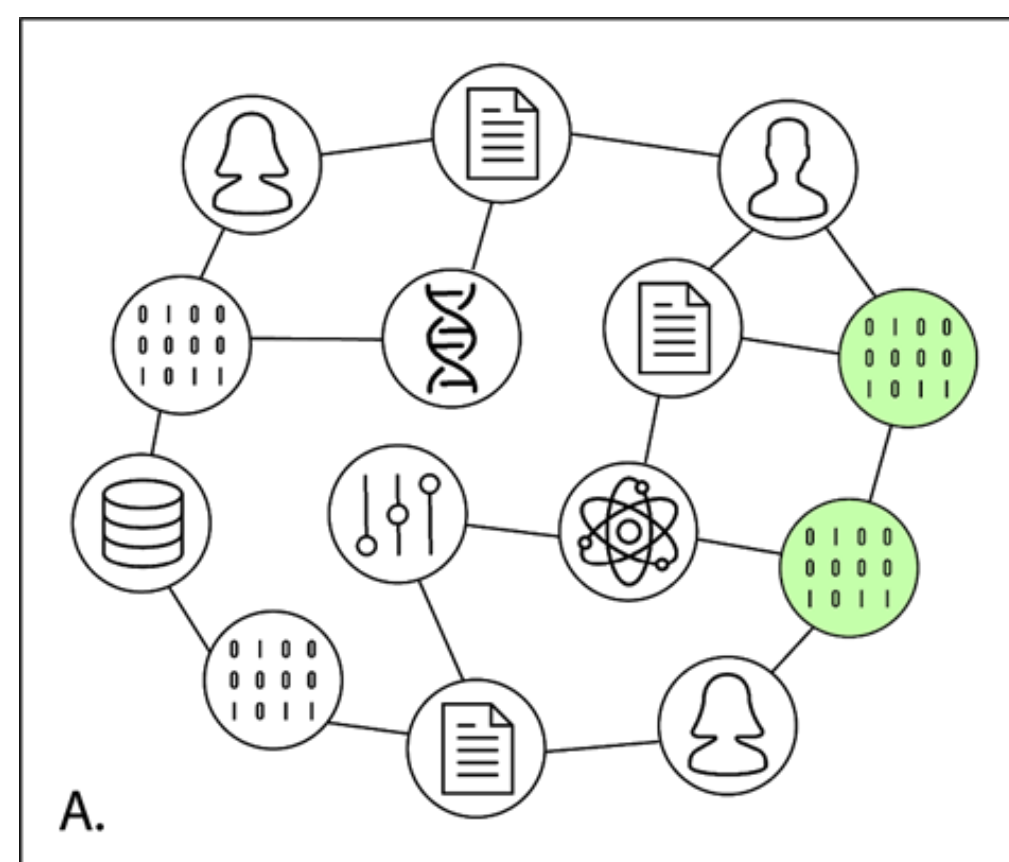


PID グラフ



DataCiteは、既存の永続的識別子を標準化された方法で相互に結びつける作業を継続している。例えば、特定の研究者、リポジトリ、機関、資金提供者を、関連するアウトプットに結びつけることで、発見やインパクト評価を行うことができる。

これらのユースケースに対処するためには、PID によって識別されるリソースとその間の結びつきを記述するためのより複雑なモデル、すなわちグラフが必要となる。グラフ理論では、PIDによって識別されるリソースはこのグラフのノードに対応し、PID間の結びつきはエッジに対応する。



People, places and things



The DataCite metadata schema supports linking people, places and things using identifiers. The API that powers the PID Graph, the graph formed by scholarly resources described by persistent identifiers (PIDs) and the connections between the FAIR Digital Objects (FDOs). The technology is powered by GraphQL, a widely adopted Open Source technology that enables queries of this graph, addressing use cases of our community in ways that were not possible before.



人、場所、もの



DataCiteメタデータスキーマは、識別子を使って人、場所、ものをリンクすることをサポートしている。永続的な識別子(PID)で記述された学術リソースによって形成されるPIDグラフと、FAIRデジタルオブジェクト(FDO)間の接続を支援するAPIである。この技術は、このグラフのクエリを可能にするオープンソース技術として広く採用されているGraphQLによって提供されており、以前は不可能だった方法によって、コミュニティのユースケースに対応している。

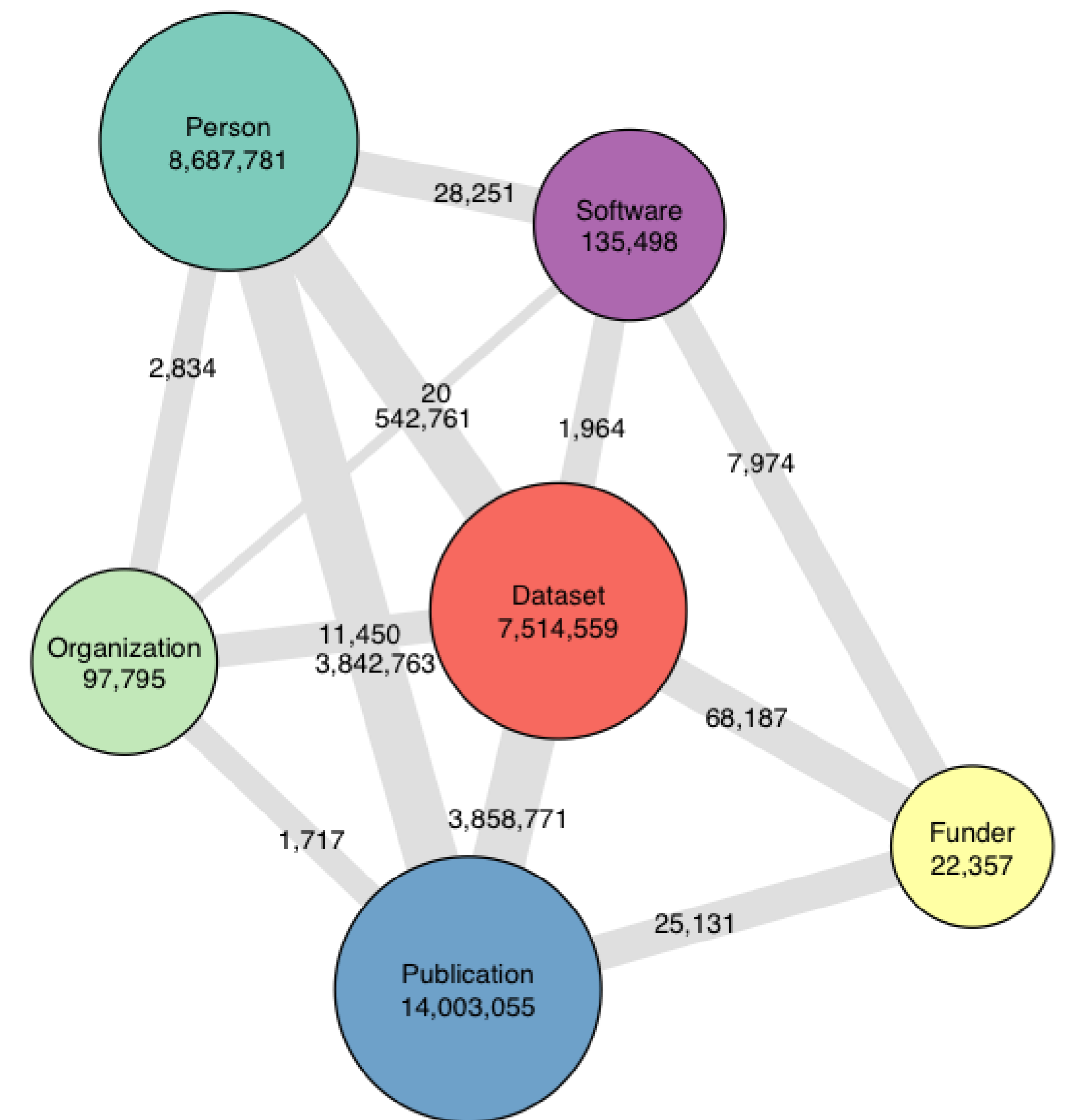


Research is already a graph



Researchers, institutions, publications, datasets, and more are interconnected.

Entities and the relationships between them form a conceptual graph of the connected research landscape.

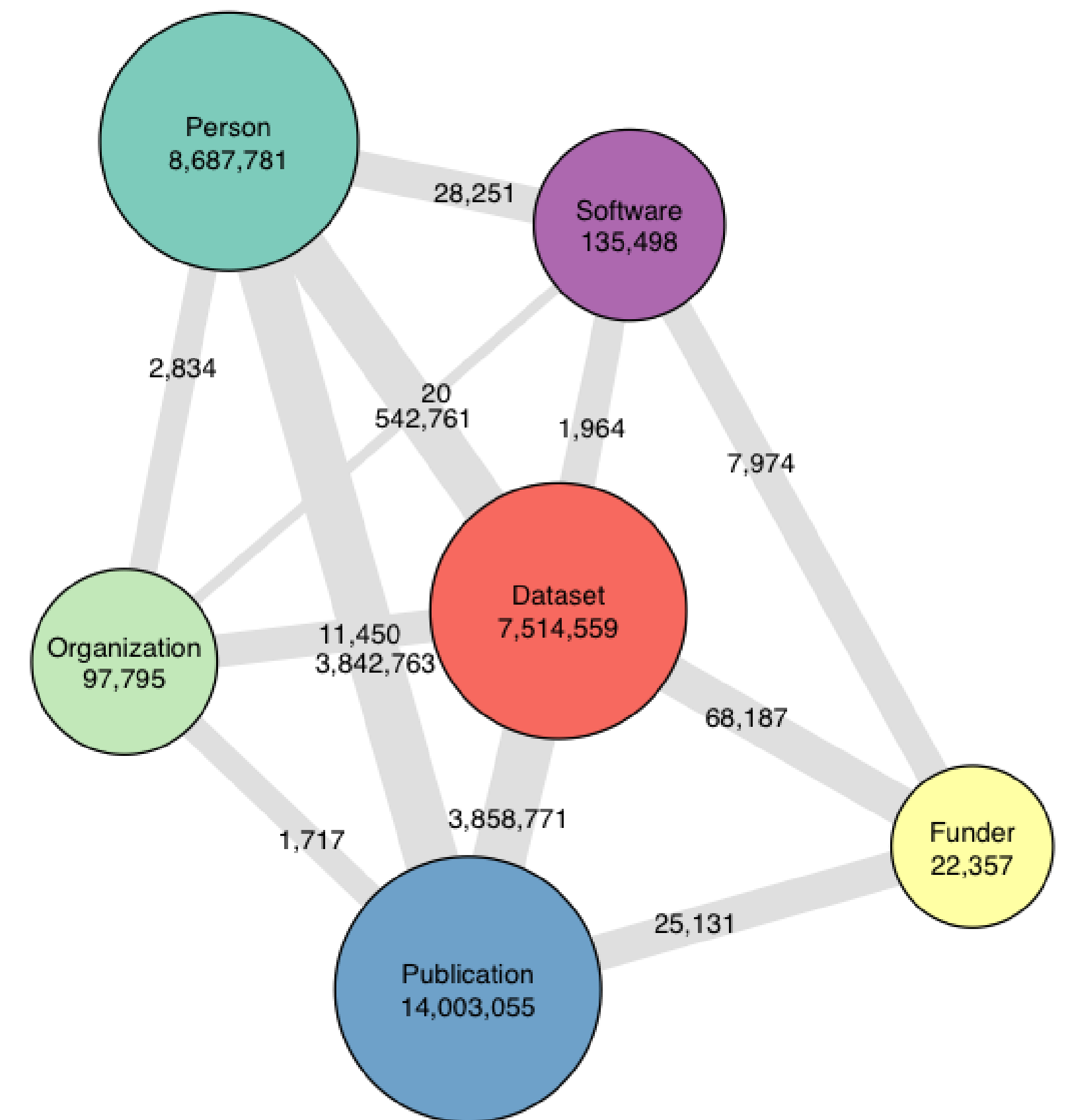


研究はすでにグラフ

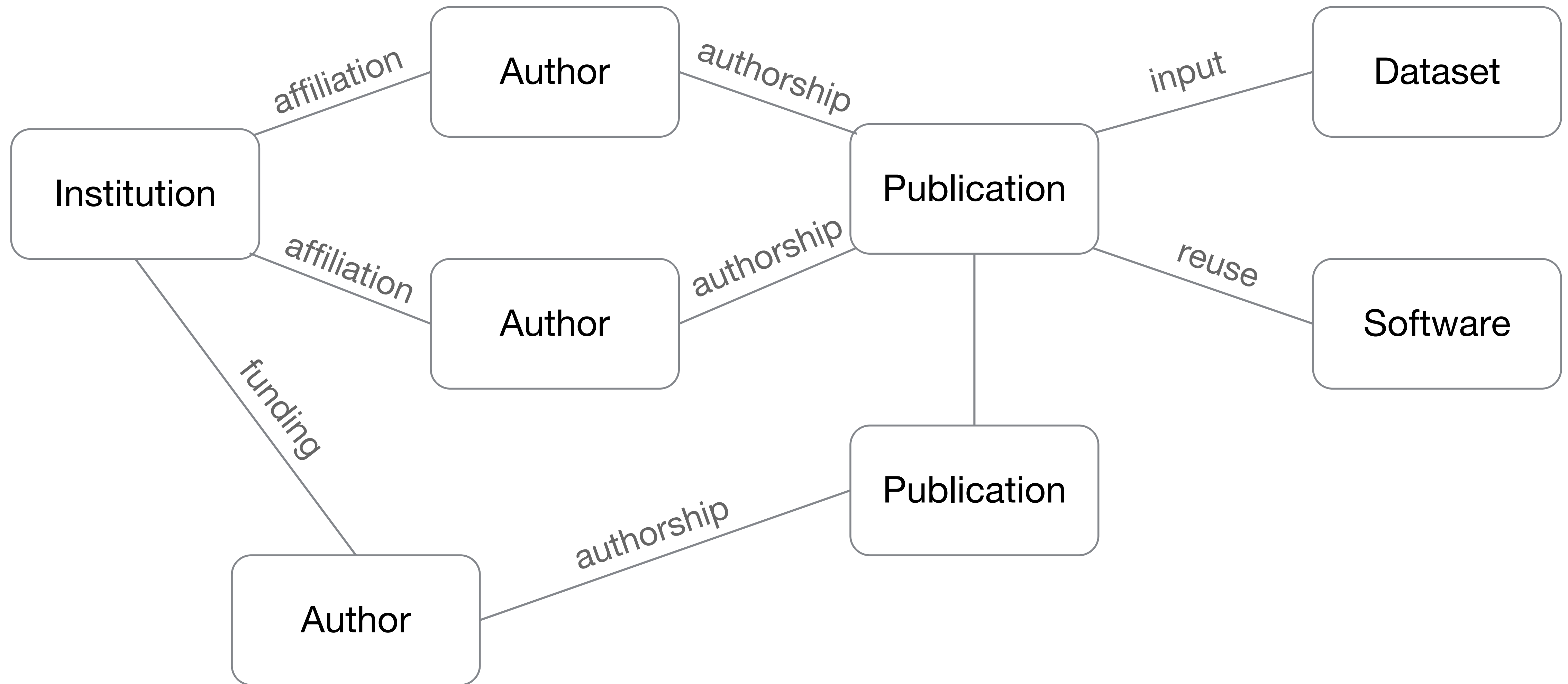


研究者、機関、出版物、データセットなどが相互に結びついている。

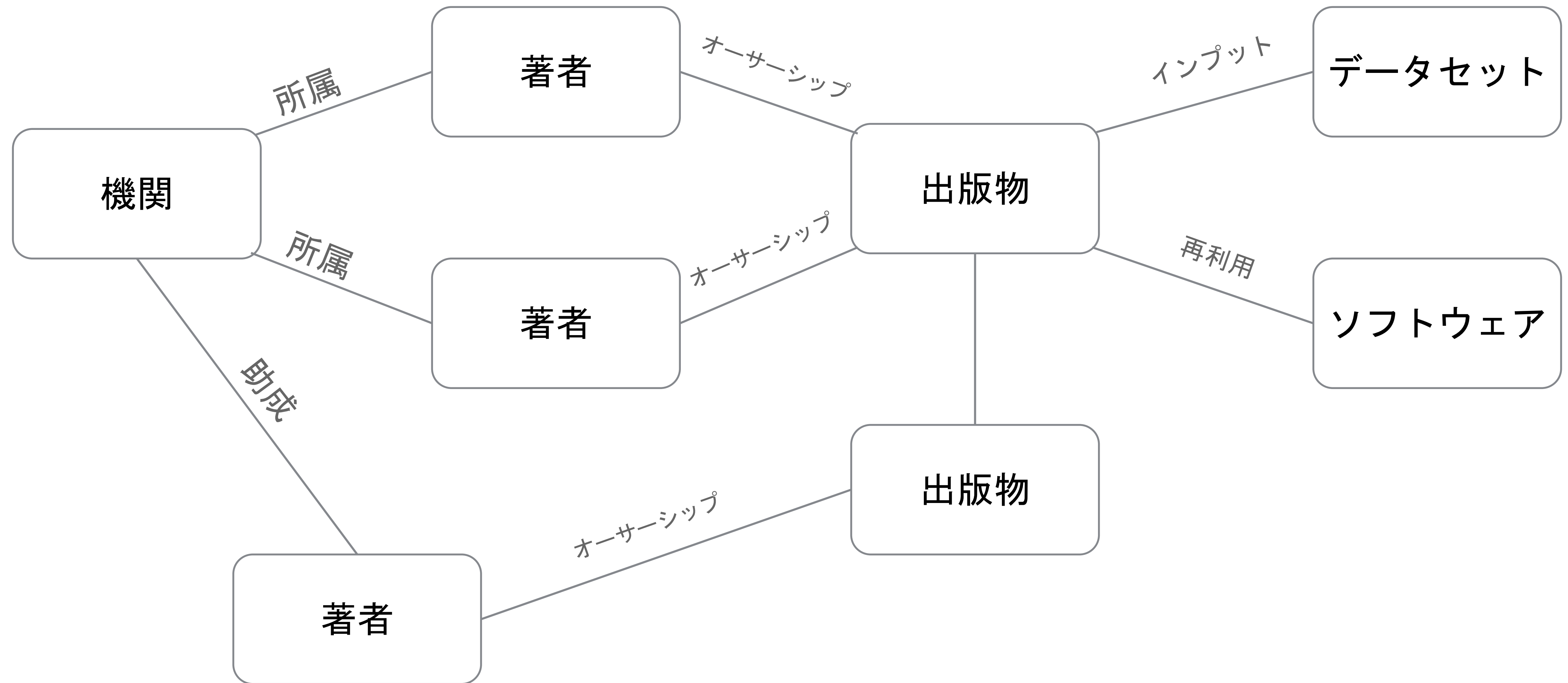
これらの関係性が、研究ランドスケープの概念的なグラフを形成している。



It could look like this...



こんな感じでは...



PIDs are the backbone of connected research



Having unique persistent identifiers for researchers and their outputs is crucial to connecting pieces of the research landscape together.

PIDs already have the potential to enable the connected research graph, but we're not yet taking full advantage of their connecting powers.

PIDは研究をつなげるためのバックボーンである



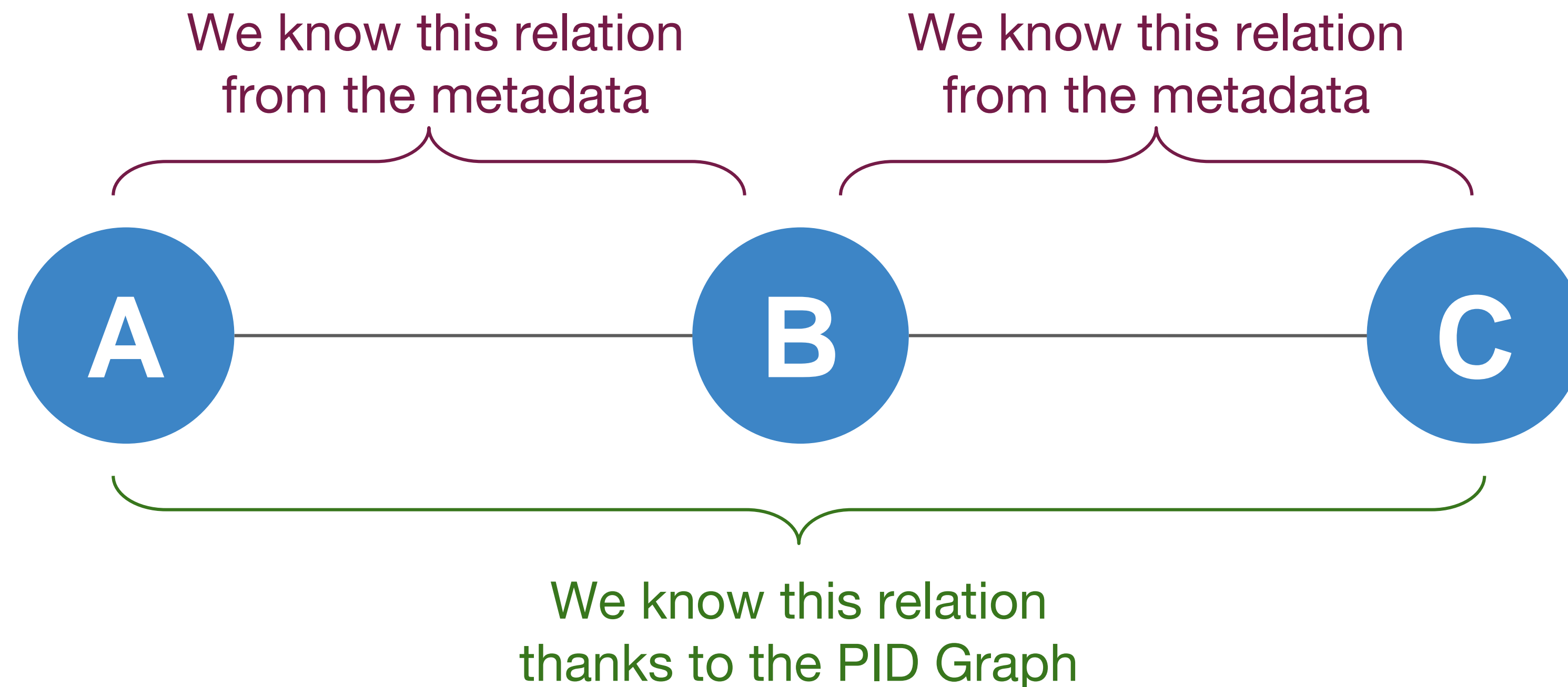
研究者とそのアウトプットに対する固有の永続的な識別子を持つことは、研究の断片をつなぎ合わせるために非常に重要である。

PIDは研究のつながりを可視化するグラフの実現可能性を既に秘めているが、我々はまだそのつながりの力を十分に活用できていない。

Enter the PID Graph



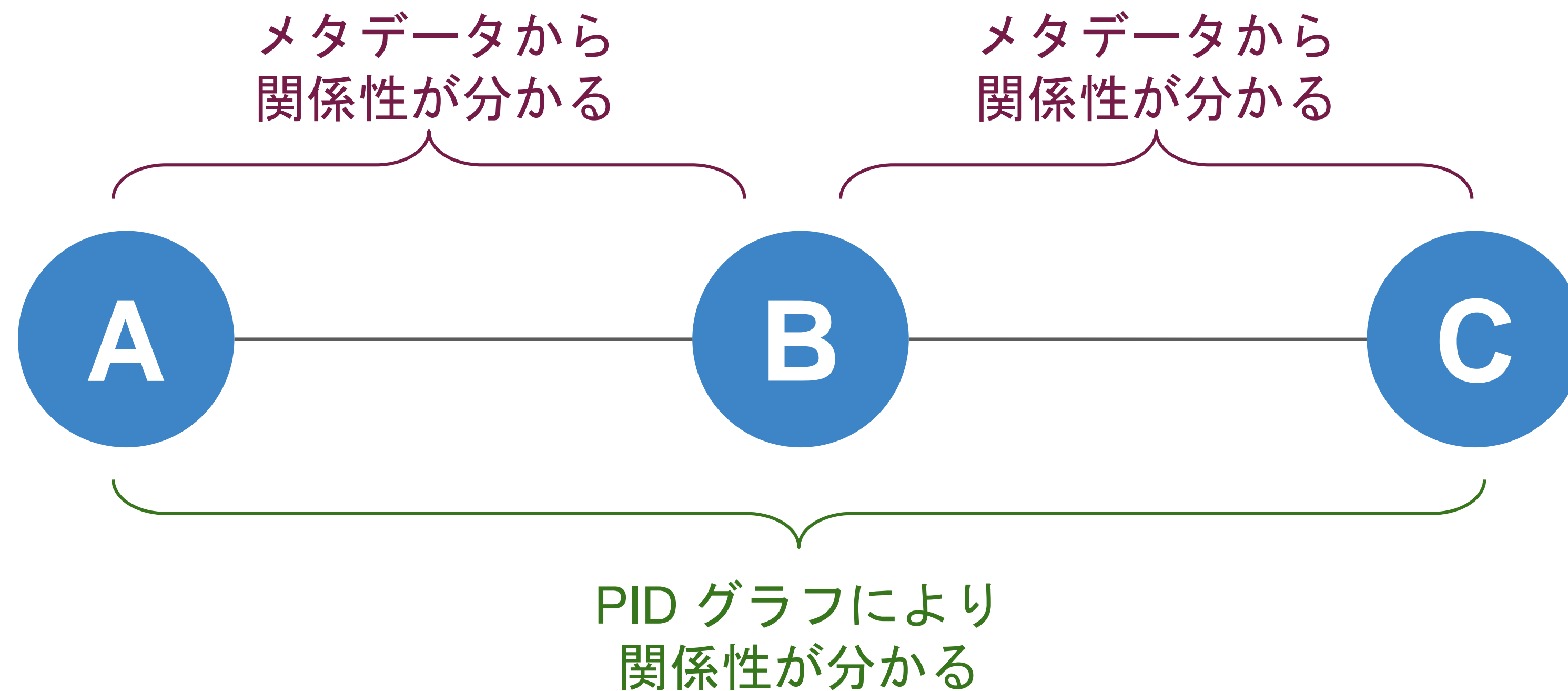
We can link PIDs together via relations in their metadata to enable the discovery of connections at least two “hops” away.



PID グラフの導入



メタデータ内の関係を介してPIDをリンクし、
少なくとも2つ先の結びつきが発見できる



DataCite GraphQL API



DataCite has developed a GraphQL API. This is the most convenient way to consume the PID Graph with DataCite metadata as a starting point.

GraphQL is a query language that's specially built for graphs. It lets you specify exactly which information you want and in what structure you'd like to receive it.

The DataCite GraphQL API is currently in pre-release, with a stable release planned for around RDA in March.

DataCite GraphQL API



DataCiteはGraphQL APIを開発した。これは、DataCiteのメタデータを起点に、PIDグラフを使うための最も便利な方法である。

GraphQLはグラフ用に特別に作られたクエリ言語である。どの情報をどのような構造で受け取るかを正確に指定することができる。

DataCite GraphQL APIは現在プレリリース中で、3月のRDA開催時期には、安定版のリリースが予定されている。

GraphQL Endpoint

https://api.datacite.org/graphql

Method

POST

Edit HTTP Headers

GraphiQL



Prettify

History

< Docs

```
1 {
2   publications(query: "creators.name:dasler") {
3     id
4     titles {
5       title
6     }
7     descriptions {
8       description
9     }
10    creators {
11      id
12      name
13      familyName
14    }
15    fundingReferences {
16      funderIdentifier
17      funderName
18      awardTitle
19      awardNumber
20    }
21  }
22 }
23
```

```
{
  "data": {
    "publications": [
      {
        "id": "https://doi.org/10.5281/zenodo.1064000",
        "titles": [
          {
            "title": "Pid Service Adoption"
          }
        ],
        "descriptions": [
          {
            "description": "This presentation describes how the uptake of persistent identifiers can be measured and gives an overview of the main results of the ORCID adoption study."
          }
        ],
        "creators": [
          {
            "id": null,
            "name": "Dasler, Robin",
            "familyName": "Dasler"
          }
        ]
      }
    ]
  }
}
```

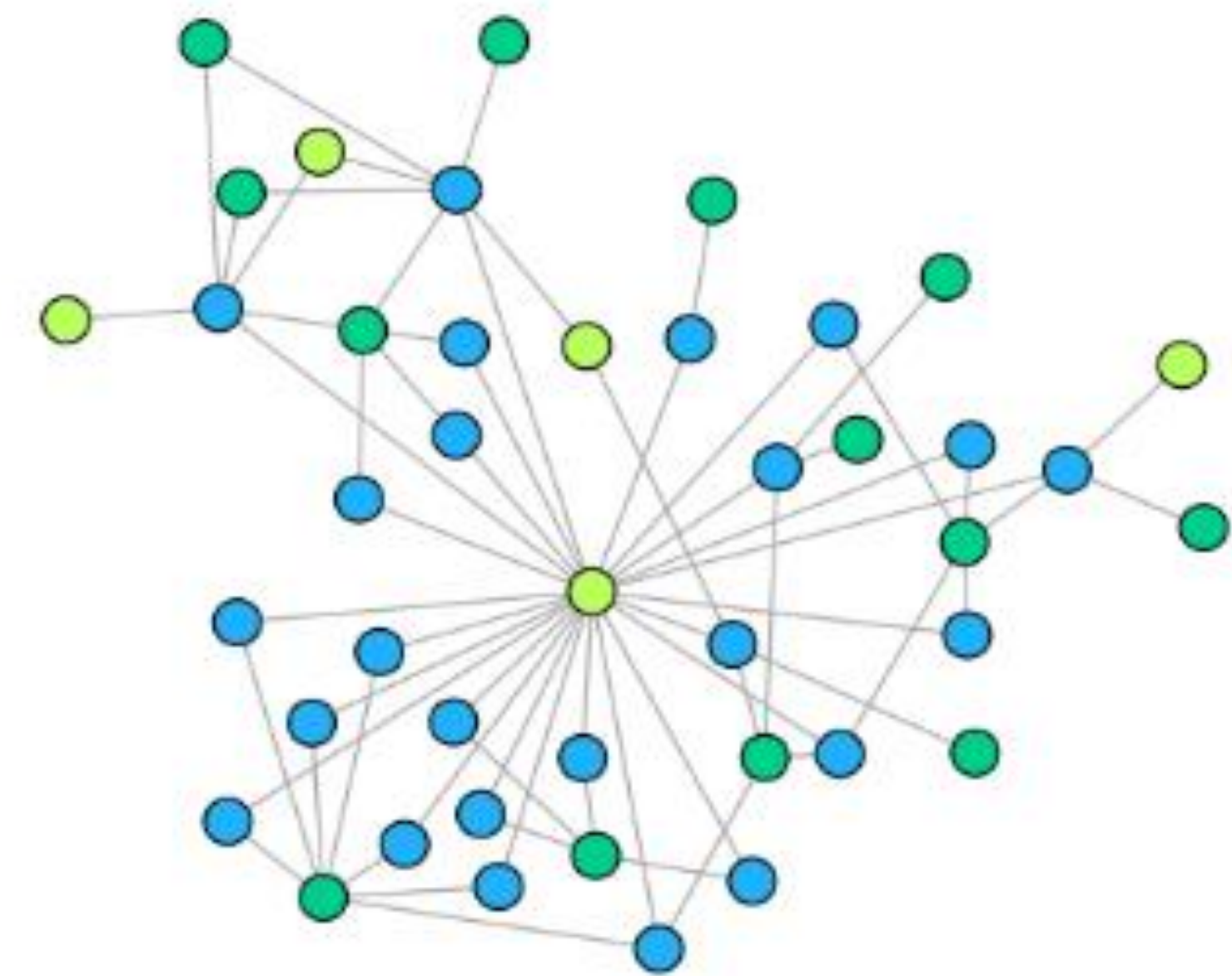
Surfacing connections



Graph of all the publications associated with DataCite, plus all the researchers and organizations associated with those publications.

Answers the question :

With whom has DataCite collaborated on their publications?



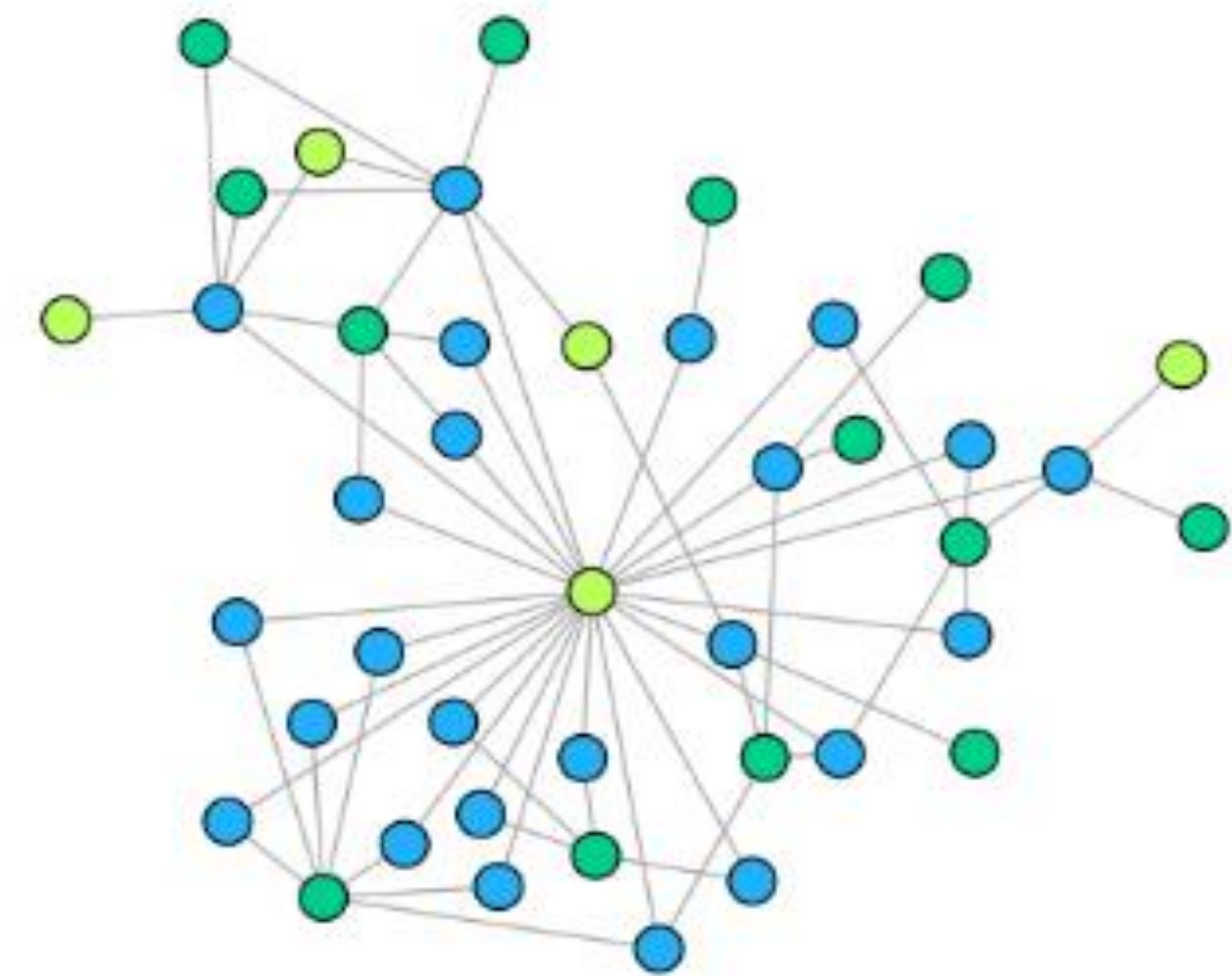
表面的な接続



DataCiteに関連するすべての出版物、それらの出版物に関連するすべての研究者と組織のグラフ。

質問にお答えしましょう。

DataCiteは出版物において、誰とコラボレーションしているか？



● 機関
● 出版物
● 研究者

PRETTIFY

HISTORY

 <https://api.datacite.org/graphql>

COPY CURL

1



DOCS

SCHEMA

Hit the Play Button to
get a response here

DataCite Commons: leveraging the PID Graph



<https://ror.org/0112mx960>

Tokyo Institute of Technology

Tokyo Tech, Tōkyō Kōgyō Daigaku, Tokodai, TIT

Links

[Homepage](#)

[Wikipedia](#)

Other Identifiers

GRID [grid.32197.3e](#)

Crossref Funder ID [10.13039/501100004520](#)

ISNI [0000000121792105](#)

Wikidata [Q587326](#)

Wikidata [Q39664847](#)

Wikidata [Q39756382](#)

[Japan](#)

[Education](#)

<https://ror.org/0112mx960>

DataCite Commons: PIDグラフの活用



<https://ror.org/0112mx960>

Tokyo Institute of Technology

Tokyo Tech, Tōkyō Kōgyō Daigaku, Tokodai, TIT

Links

[Homepage](#)

[Wikipedia](#)

Other Identifiers

GRID [grid.32197.3e](#)

Crossref Funder ID [10.13039/501100004520](#)

ISNI [0000000121792105](#)

Wikidata [Q587326](#)

Wikidata [Q39664847](#)

Wikidata [Q39756382](#)

[Japan](#)

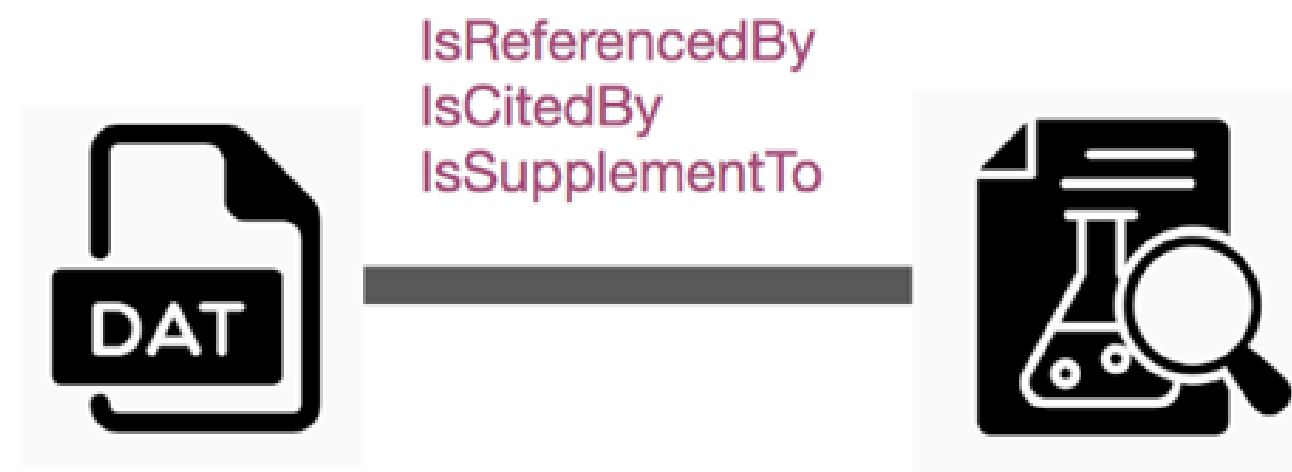
[Education](#)

<https://ror.org/0112mx960>

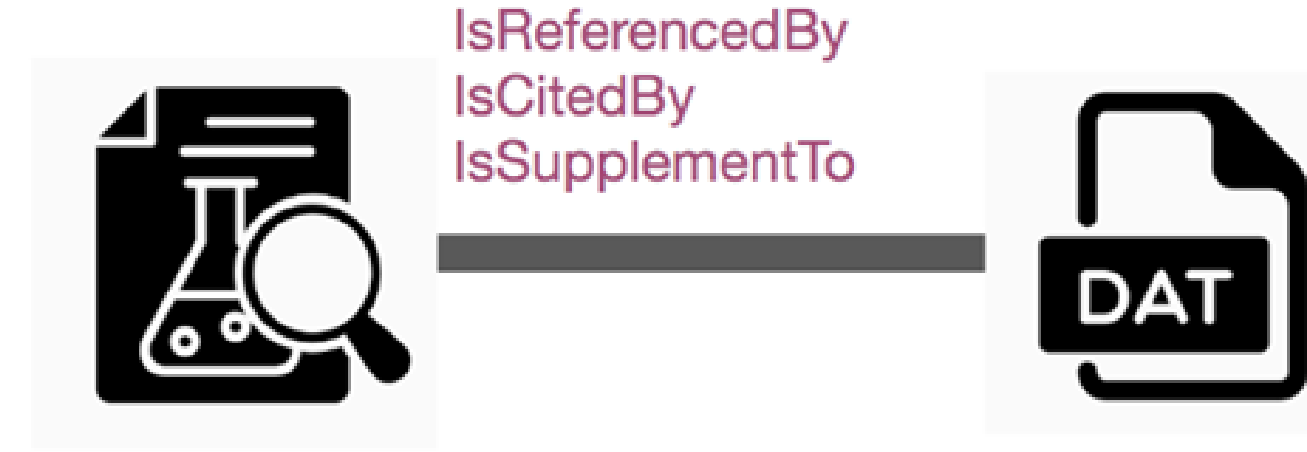
PID Relations



Citations



References



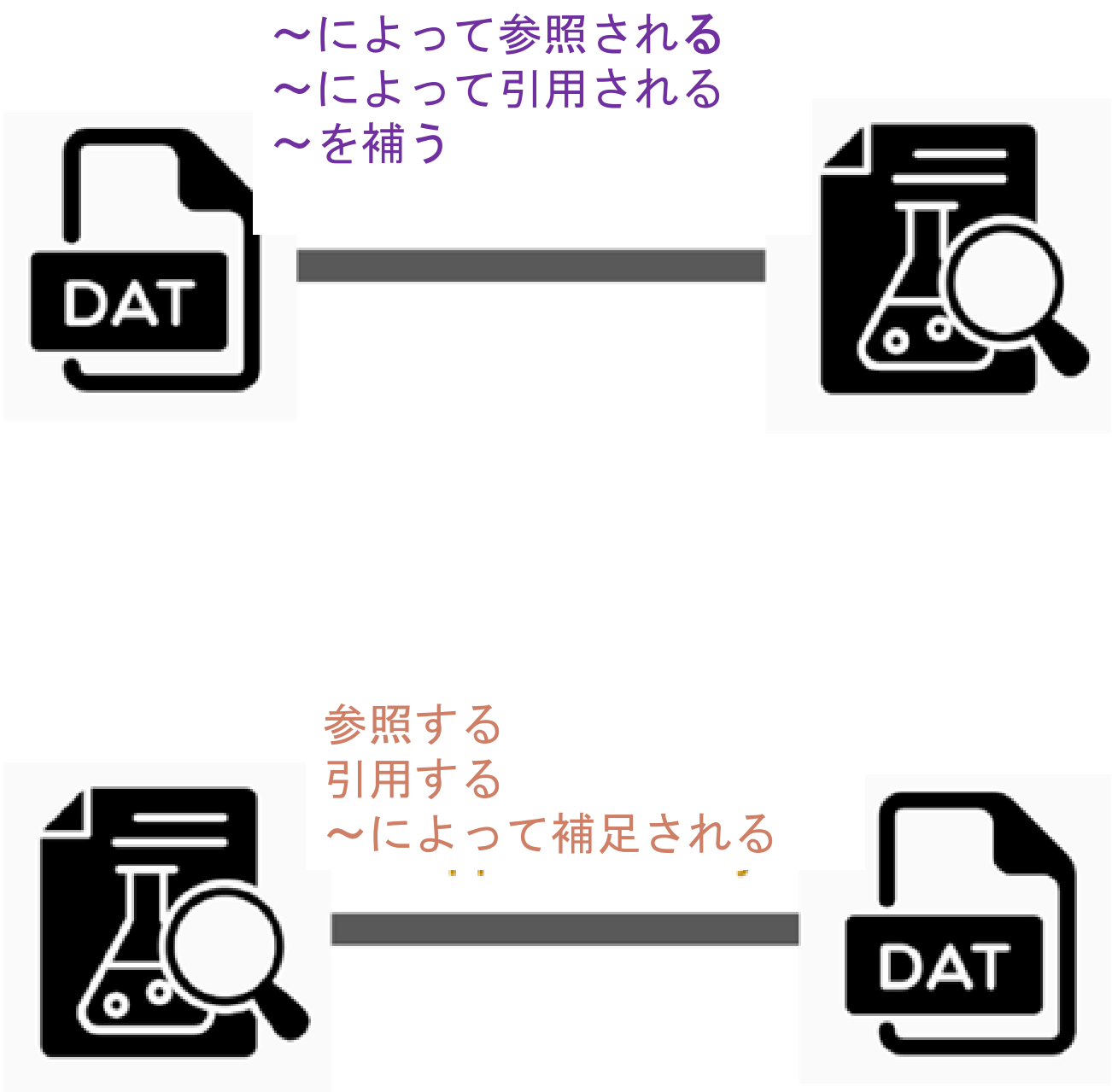
Relations

IsContinuedBy
Continues
IsDescribedBy
Describes
HasMetadata
IsMetadataFor
HasVersion
IsVersionOf
IsNewVersionOf
IsPreviousVersionOf
IsPartOf
HasPart
IsDocumentedBy
Documents
IsCompiledBy
Compiles
IsVariantFormOf
IsOriginalFormOf
IsIdenticalTo
IsReviewedBy
Reviews
IsDerivedFrom
IsSourceOf
IsRequiredBy
Requires
IsObsoletedBy
Obsoletes

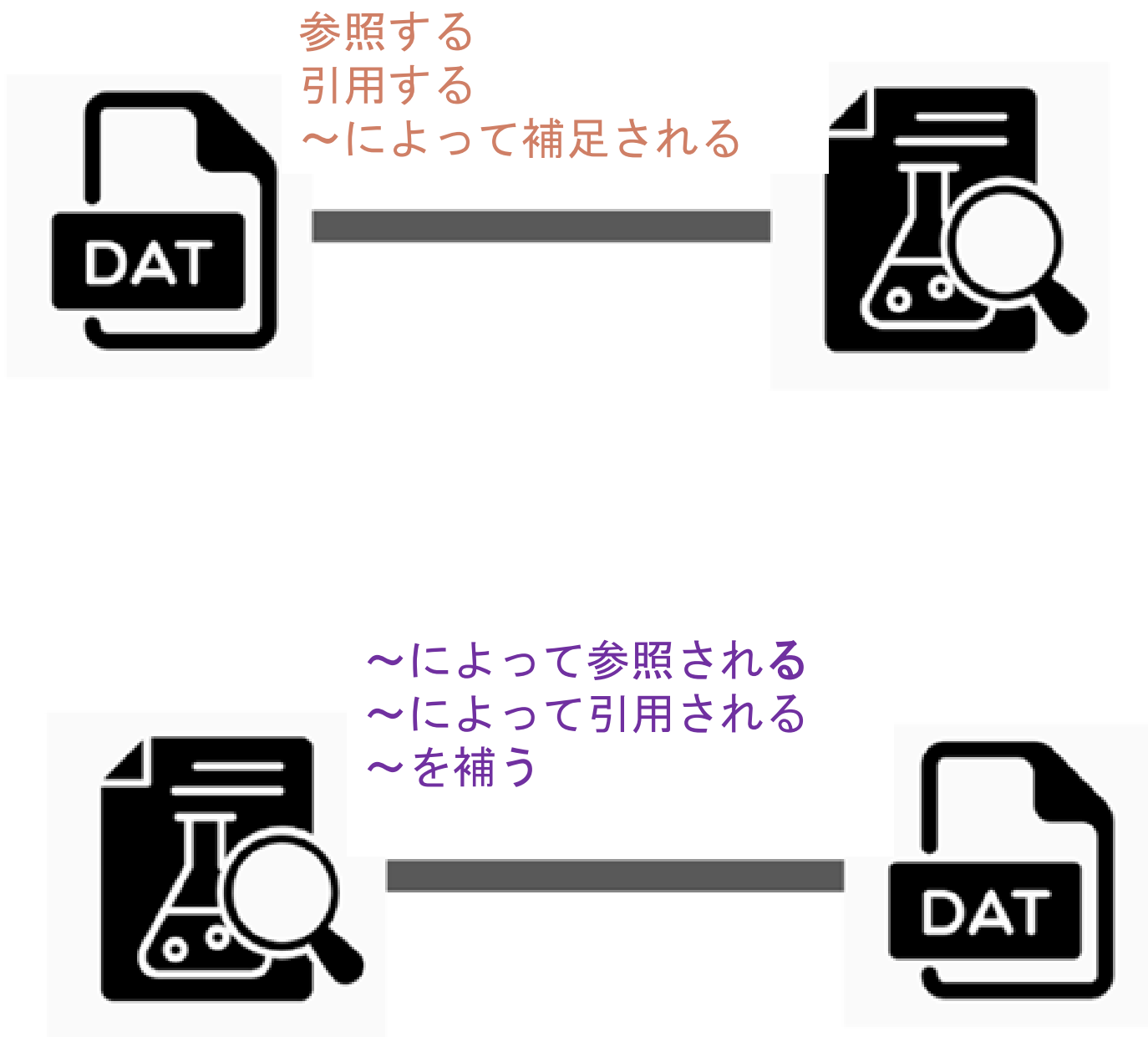
PIDの関係性



引用



参照



関係性

IsContinuedBy
Continues
IsDescribedBy
Describes
HasMetadata
IsMetadataFor
HasVersion
IsVersionOf
IsNewVersionOf
IsPreviousVersionOf
IsPartOf
HasPart
IsDocumentedBy
Documents
IsCompiledBy
Compiles
IsVariantFormOf
IsOriginalFormOf
IsIdenticalTo
IsReviewedBy
Reviews
IsDerivedFrom
IsSourceOf
IsRequiredBy
Requires
IsObsoletedBy
Obsoletes

Citations & Usage



Data from: Impact of negative frequency-dependent selection on mating pattern and genetic structure: a comparative analysis of the S-locus and nuclear SSR loci in *Prunus lannesiana* var. *speciosa*

Kato Shuri, Teruyoshi Nagamitsu, Hiroyoshi Iwata, Yoshihiko Tsumura, Yuzuru Mukai, K Michiharu, K Saika & K Junko

Version 1 of Dataset published 2012 in [DRYAD](#)

Mating processes of local demes and spatial genetic structure of island populations at the self-incompatibility (S-) locus under negative frequency-dependent selection (NFDS) were evaluated in *Prunus lannesiana* var. *speciosa* in comparison with nuclear simple sequence repeat (SSR) loci that seemed to be evolutionarily neutral. Our observations of local mating patterns indicated that male-female pair fecundity was influenced by not only self-incompatibility, but also various factors such as kinship, pollen production and flowering synchrony. In spite of the mating bias caused by these factors, the NFDS effect on changes in allele frequencies from potential mates to mating pollen was detected at the S-locus but not at the SSR loci although the changes from adult to juvenile cohorts were not apparent at any loci. Genetic differentiation and isolation-by-distance over various spatial scales were smaller at the S-locus than at the SSR loci, as expected under the NFDS. All allele sharing distributions among the populations also had a unimodal pattern at the S-locus, indicating the NFDS effect except for alleles unique to individual populations probably due to isolation among islands, although this pattern was not exhibited by the SSR loci. Our results suggest that the NFDS at the S-locus has an impact on both the mating patterns and the genetic structure in the *P. lannesiana* populations studied.

DOI registered April 17, 2012 via DataCite.



1 Citation 99 Views 16 Downloads

[Dataset](#)

[English](#)

<https://doi.org/10.5061/dryad.7c425>



Data from: Impact of negative frequency-dependent selection on mating pattern and genetic structure: a comparative analysis of the S-locus and nuclear SSR loci in *Prunus lannesiana* var. *speciosa*

Kato Shuri, Teruyoshi Nagamitsu, Hiroyoshi Iwata, Yoshihiko Tsumura, Yuzuru Mukai, K Michiharu, K Saika & K Junko

Version 1 of Dataset published 2012 in [DRYAD](#)

Mating processes of local demes and spatial genetic structure of island populations at the self-incompatibility (S-) locus under negative frequency-dependent selection (NFDS) were evaluated in *Prunus lannesiana* var. *speciosa* in comparison with nuclear simple sequence repeat (SSR) loci that seemed to be evolutionarily neutral. Our observations of local mating patterns indicated that male-female pair fecundity was influenced by not only self-incompatibility, but also various factors such as kinship, pollen production and flowering synchrony. In spite of the mating bias caused by these factors, the NFDS effect on changes in allele frequencies from potential mates to mating pollen was detected at the S-locus but not at the SSR loci although the changes from adult to juvenile cohorts were not apparent at any loci. Genetic differentiation and isolation-by-distance over various spatial scales were smaller at the S-locus than at the SSR loci, as expected under the NFDS. All allele sharing distributions among the populations also had a unimodal pattern at the S-locus, indicating the NFDS effect except for alleles unique to individual populations probably due to isolation among islands, although this pattern was not exhibited by the SSR loci. Our results suggest that the NFDS at the S-locus has an impact on both the mating patterns and the genetic structure in the *P. lannesiana* populations studied.

DOI registered April 17, 2012 via DataCite.



1 Citation 99 Views 16 Downloads

[Dataset](#)

[English](#)

<https://doi.org/10.5061/dryad.7c425>

Creators, Citations, Usage



Creators

Kato Shuri
Forestry and Forest Products Research
Institute

Yoshihiko Tsumura
Forestry and Forest Products Research
Institute

K Saika
Tokyo Institute of Technology

Teruyoshi Nagamitsu
Forestry and Forest Products Research
Institute

Yuzuru Mukai
Gifu University

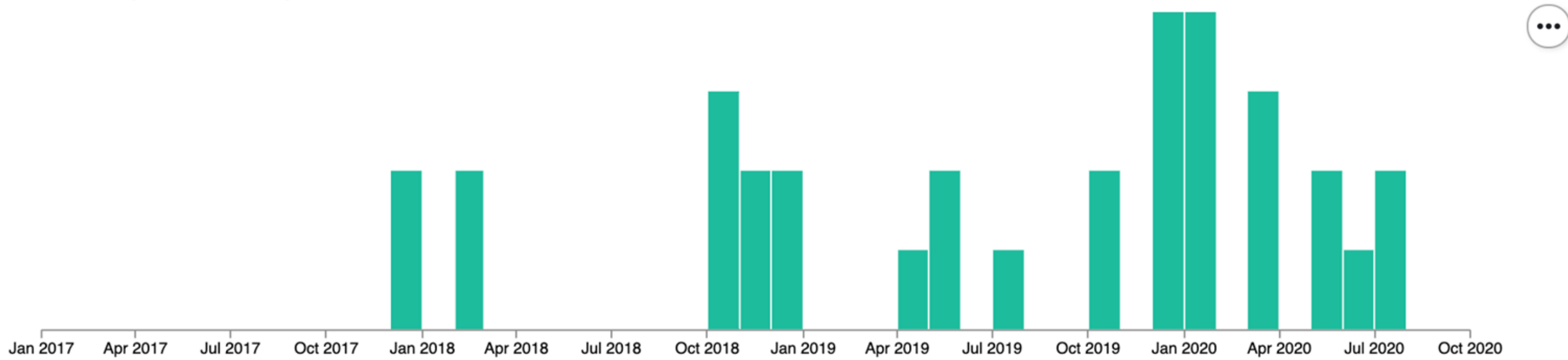
K Junko
Gunma University

Hiroyoshi Iwata
University of Tokyo

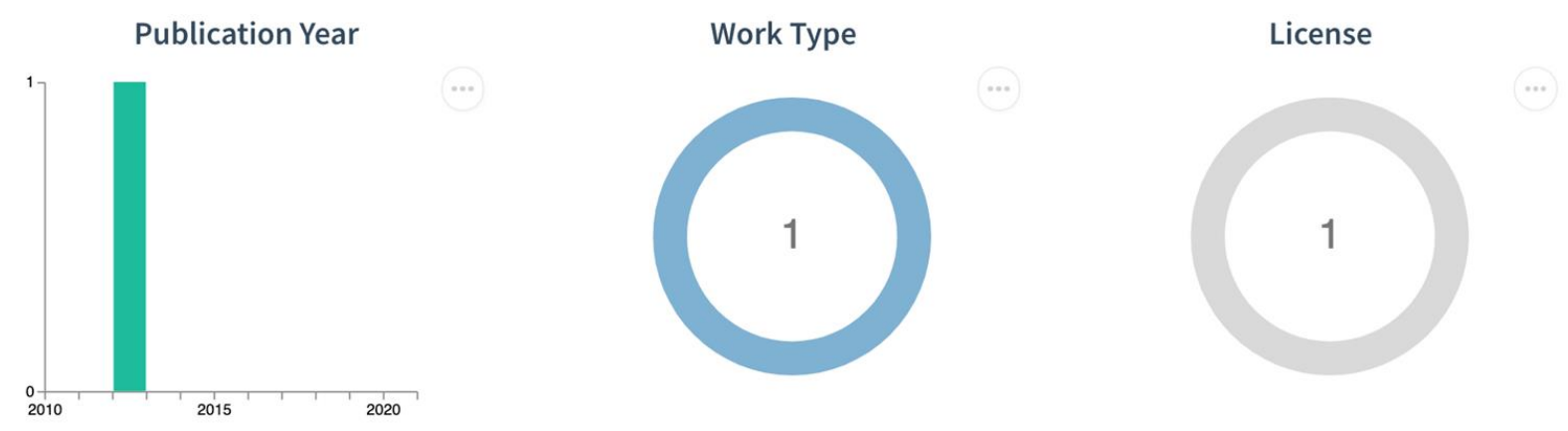
K Michiharu
Kyoto University

99 Views 16 Views

99 views reported since publication in 2012.



1 Reference 1 Citation



Impact of negative frequency-dependent selection on mating pattern and genetic structure: a comparative analysis of the S-locus and nuclear SSR loci in *Prunus lannesiana* var. *speciosa*

K Shuri, K Saika, K Junko, K Michiharu, T Nagamitsu, H Iwata, Y Tsumura & Y Mukai
Journal Article published 2012 in [Heredity](#)

DOI registered via Crossref.

1 Citation

[Journal Article](#)

<https://doi.org/10.1038/hdy.2012.29>

作者、引用、利用



Creators

Kato Shuri
Forestry and Forest Products Research
Institute

Yoshihiko Tsumura
Forestry and Forest Products Research
Institute

K Saika
Tokyo Institute of Technology

Teruyoshi Nagamitsu
Forestry and Forest Products Research
Institute

Yuzuru Mukai
Gifu University

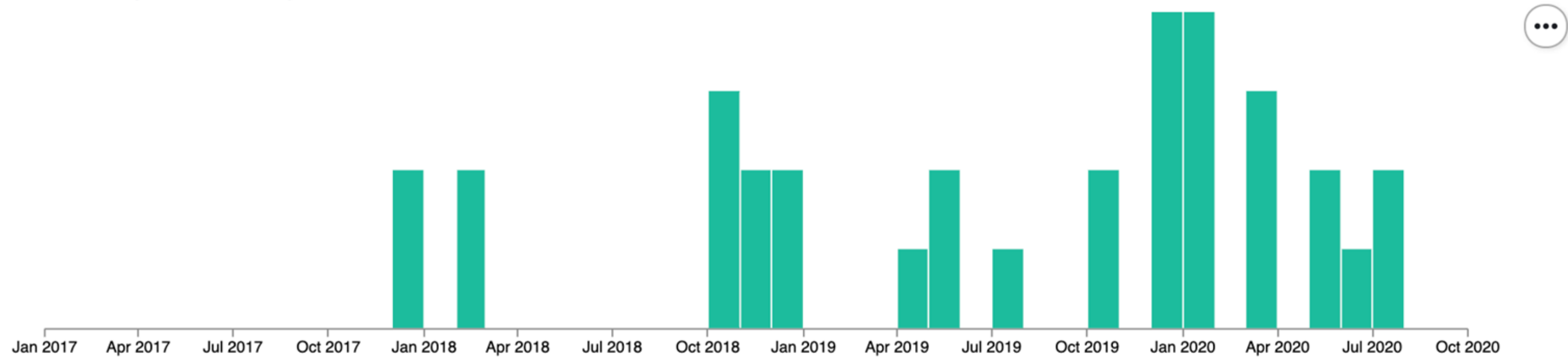
K Junko
Gunma University

Hiroyoshi Iwata
University of Tokyo

K Michiharu
Kyoto University

99 Views 16 Views

99 views reported since publication in 2012.



1 Reference 1 Citation



Impact of negative frequency-dependent selection on mating pattern and genetic structure: a comparative analysis of the S-locus and nuclear SSR loci in *Prunus lannesiana* var. *speciosa*

K Shuri, K Saika, K Junko, K Michiharu, T Nagamitsu, H Iwata, Y Tsumura & Y Mukai
Journal Article published 2012 in [Heredity](#)

DOI registered via Crossref.

1 Citation

[Journal Article](#)

<https://doi.org/10.1038/hdy.2012.29>

Metadata to improve reuse & discovery



- **Rights** - The Rights field is where you can include information about the data's content license (like CC0). When this information is in the metadata for the dataset, it helps researchers to quickly see at a glance whether the data being described is something they'll be able to use. This information is also important for harvesters who are trying to assemble lists of readily available datasets.
- **Description** - The Description field may at first seem not particularly critical, but abstracts and other descriptive information is often mined for emerging trends. In a case like the current epidemic, where "COVID-19" might not be a typical controlled vocabulary term, it may be possible to surface articles on this topic by mining abstracts. Note that our schema documentation also recommends providing an English-language abstract to enable wider dissemination.

再利用と発見を改善するためのメタデータ



- **Rights（権利）** - Rightsフィールドには、データのコンテンツライセンス（CC0のようなもの）に関する情報を含めることができる。この情報がデータセットのメタデータに含まれていると、研究者は、記述されているデータが利用可能なものであるかどうかを一目で確認することができる。この情報は、すぐに利用可能なデータセットのリストを集めようとしている情報収集者にとっても重要である。
- **Description（説明）** - Descriptionフィールドは、最初は特に重要ではないように見えるかもしれないが、抄録やその他の説明的な情報は、しばしば新たな傾向を探るために収集される。昨今の伝染病のようなケースでは、「COVID-19」が典型的な統制語でなくても、抄録をマイニングすることで、このトピックに関する記事を表面化することができるかもしれない。私達の概要説明資料では、より広く普及させるために英語の抄録を提供することを推奨している。

Metadata to improve reuse & discovery (cont.)



- **RelatedIdentifier** - This one may go without saying. RelatedIdentifiers are always important for linking resources together. This is the information that DataCite uses for our citation counts and to feed the PID Graph. Having this RelatedIdentifier information is also helpful for following a trail of research from dataset to article to author to institution and so on.
- **Subject** - The Subject field is where you can include information about subjects that are relevant for the item the DOI points to. Under the current implementation of the DataCite Metadata Schema, this field is free text, so you can add any keywords that might be relevant to the item, whether or not they're part of a controlled vocabulary. If you would like to follow a scheme of vocabulary, you're not restricted in the subject scheme you can specify. It can be helpful to comply with community-specific standards, such as using MeSH subject headings for medical topics or using DFG subjects for German-funded research, for example.

再利用と発見を改善するためのメタデータ(続)



- **RelatedIdentifier (関連識別子)** - これは言わずもがなかかもしれない。RelatedIdentifierは、リソースを相互にリンクさせるために常に重要である。これはDataCiteが引用数カウントやPIDグラフの提供に使用している情報である。この RelatedIdentifierの 情報を持っていると、データセットから論文、著者、機関などの研究の軌跡をたどるのにも役立つ。
- **Subject (主題)** - Subjectフィールドには、DOIが指し示す項目に関連する主題に関する情報を含めることができる。DataCiteメタデータスキーマの昨今の実装では、このフィールドはフリーテキストなので、統制語の一部であるかどうかに関わらず、項目に関連する可能性のあるキーワードを追加することができる。語彙のスキームに従いたい場合は、指定できる主題スキームに制限はない。例えば、医学的なトピックにはMeSHの主題見出しを使用したり、ドイツの資金提供を受けた研究にはDFGの主題を使用したりするなど、コミュニティ固有の基準に準拠するのに役立つことがある。

Community of Practice



Our Domain

- Connecting research, identifying knowledge.

Our Community

- Global research community

Our Practice

- Persistent identifiers for research.
- Connect PIDs and related metadata with the broader research ecosystem.
- Create, find, cite, connect, and use research.



実践のコミュニティ



私たちのドメイン

- 研究をつなぐ、知識を特定する

私たちのコミュニティ

- グローバルな研究コミュニティ

業務内容

- 研究のための永続的な識別子
- PIDと関連するメタデータをより広範な研究エコシステムに接続
- 研究の作成、検索、引用、接続、利用



PID Graph & Open Science



- The PID Graph enables *processes that enable the creation, contribution, discovery and reuse of research knowledge more reliably, effectively and equitably*
- Together as a community we can further our Open Science efforts



PID グラフ & オープンサイエンス



- PIDグラフは研究知識の創造、貢献、発見、再利用をより確実に、効果的に、そして公平に実現するプロセスを可能にする。
- 我々はコミュニティとして、オープンサイエンスの取り組みをさらに発展させることができる。



Get involved



- Let us know if you are interested in the PID Graph, DataCite and JaLC collaborative efforts
info@datacite.org
- Try out the PID Graph playground
<https://api.datacite.org/graphql>
- Provide feedback in the PID Forum
<https://www.pidforum.org/c/pid-graph/17>

参加しよう



- PIDグラフ、DataCite、JaLCの共同作業に興味のある方は、
info@datacite.org
- PIDグラフの遊び場を試してみたい方は、
<https://api.datacite.org/graphql>
- PID フォーラムへフィードバックを提供してくださる方は、
<https://www.pidforum.org/c/pid-graph/17>

へご連絡ください。

Building communities together!

matt.buys@datacite.org
[@datacite](https://twitter.com/datacite)
datacite.org



みんなでコミュニティを作しましょう！

matt.buys@datacite.org
@datacite
datacite.org
