The PID Graph
Realizing the power of relational metadata

Matt Buys
June 15, 2021
Japan Open Science Summit
Todays research, tomorrow’s discoveries

Discoverability
- Increase (enable) open access to the results of research
- Ensure the dissemination and reuse of research

Trust
- Recognize the impact and reuse of research outputs
- Preserve the research data used in studies
- Bring more rigor to the scholarly record
It is important that the focus is not exclusively persistent identifiers (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.
Research is already a 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. This is possible using **relational metadata**.
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.

We continue to extend the PID Graph and include various other identifiers in the relational metadata.
Meaning through metadata

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

It is possible for institutions to build platforms or services that make use of the PID Graph.

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. In other words, the PID Graph allows a “known query search” and enables several downstream use cases to understand relationships between research objects (people, places and things).
DataCite Commons: Leveraging the PID Graph

https://ror.org/0112mx960

Tokyo Institute of Technology
Tokyo Tech, Tōkyō Kögyō 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
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
PID Relations

**Citations**
- IsReferencedBy
- IsCitedBy
- IsSupplementedBy

**References**
- References
- Cites
- IsSupplementedBy

**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
Creators, citations and usage

Creators

Kato Shuri
Forestry and Forest Products Research Institute

Teruyoshi Nagamitsu
Forestry and Forest Products Research Institute

Hiroyoshi Iwata
University of Tokyo

Yoshihiko Tsumura
Forestry and Forest Products Research Institute

Yuzuru Mukai
Gifu University

K Michiharu
Kyoto University

K Saika
Tokyo Institute of Technology

K Junko
Gunma University

99 Views

16 Views

99 views reported since publication in 2012.

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 2022 in

DOI registered via Crossref:

https://doi.org/10.1038/s41592-022-10903-9

1 Citation

1 License
Metadata to improve discoverability and trust

- **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.
Metadata to improve discoverability and trust

- **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.
Culture change

- Make it required
- Make it rewarding
- Make it normative
- Make it easy
- Make it possible

Source: https://www.cos.io/blog/strategy-for-culture-change
Collaboration across key stakeholders

- **Focus on improving scholarly infrastructure through use cases.** The collaborative activities should ensure that we continue to improve scholarly infrastructure through addressing use cases, telling stories and showing the benefit to all involved.

- **Fostering open and inclusive relationships.** Our partnerships should always ensure that we create an open and inclusive environment for the community to participate.

- **Building a trusted ecosystem.** Trust is an integral component for community culture change.
Get involved!

Email us: matt.buys@datacite.org  
Follow us: @datacite  
Talk to us: pidforum.org  
Read about us: datacite.org  
Get support: support.datacite.org support@datacite.org