

**Une remarquablement
brève introduction au
contexte, aux
organisations, aux
bibliothèques et à
l'infrastructure ouverte
des identifiants pérennes**

Mike Nason

Bibliothécaire de la communication et de l'édition scientifique, Bibliothèques de
l'UNB | Agent de liaison pour Crossref et les métadonnées, PKP

~~Une remarquablement~~
~~brève introduction~~ au
contexte, aux
organisations, aux
bibliothèques et à
l'infrastructure ouverte
des identifiants pérennes

Mike Nason

Bibliothécaire de la communication et de l'édition scientifique, Bibliothèques de
l'UNB | Agent de liaison pour Crossref et les métadonnées, PKP

**Prendre notre temps pour
expliquer** le contexte, les
organisations, les
bibliothèques et
l'infrastructure ouverte
des identifiants pérennes

Mike Nason

Bibliothécaire de la communication et de l'édition scientifique, Bibliothèques de
l'UNB | Agent de liaison pour Crossref et les métadonnées, PKP

**Avant de
poursuivre, je dois
clarifier une
chose...**

**Les PID sont dans
l'eau potable de
l'édition scientifique.**

**Examinons ce que
nous savons.**

Les PID sont des identifiants uniques que nous attribuons à un nombre grandissant d'éléments :

- **Établissements**
- **Jeux de données**
- **Personnes**
- **Organisations**
- **Articles**
- **Monographies**
- **Séries**

Les PID sont des identifiants uniques que nous attribuons à un nombre grandissant d'éléments :

- Établissements
- Jeux de données
- Personnes
- Organisations
- Articles
- Monographies
- Séries

Vous avez peut-être constaté que les bibliothèques interviennent *de plus en plus* dans les éléments suivants :

- Édition numérique
- Profils des chercheurs
- Données de recherche
- Systèmes de GIR
- Dépôts
- Bibliométrie/collections
- Recherche ouverte

Les PID **pourraient** ou **devraient** simplifier la découverte de documents ou de recherche.

Les PID sont excellents pour la clarification et la cohérence des métadonnées, car :

- **Les noms ne sont pas uniques.**
- **Les noms ne suivent pas les règles.**
- **Les URL changent.**
- **Les lieux, les personnes, les établissements, etc. sont identifiés de multiples façons .**

L'utilisation judiciaire des
PID (et leur adoption
généralisée) pourrait faire
**gagner beaucoup de
temps** .

Les PID sont liés à des agences d'enregistrement qui collectent et distribuent les métadonnées publiquement.

**Examinons
certaines choses
moins connues.**

«La pérennité est
purement une question de
service .»

— J. Kunze , 2013

Pérenne ~~≠~~Permanent

La production d'un DOI
n'est pas homologue à
l'enregistrement d'un DOI.

«Vous savez comment
[*produire*] un [DOI], mais,
vous ne savez pas
comment [*enregistrer*] un
[DOI]. Or, la partie la plus
importante d'un [DOI] c'est
l'[*enregistrement*].
N'importe qui peut en
[*produire*].»

– Jerry Seinfeld , 1991

Les PID n'ont pas été
conçus pour être des URL
sur mesure lisibles par des
humains.

(DOI ~~≠~~ bit.ly sophistiqués)

10.1234/ 097813rhujrho7
10.1234/ journal.24.1.0001

**Les PID n'ont pas été
conçus pour être des URL
sur mesure lisibles par des
humains.**

**Ces éléments font la même
chose!
Personne ne lit les suffixes!**

**Imaginez la possibilité
d'une coquille dans un DOI
et tout le travail pour la
corriger .**

Les PID ne sont utiles **que**
dans la mesure où leurs
métadonnées sont
enregistrées .

Médiocre à l'entrée.
Médiocre à la sortie .

Les agences
d'enregistrement utilisent
différents schémas de
métadonnées dont la
compatibilité varie .

...

On n'a pas besoin
d'attribuer des PID à
tout forcément; **on ne
devrait pas non plus** .

Nous devons nous *détendre* .

Cependant ! Les PID
peuvent être attribués à
de nombreuses choses en
plus des articles de revue
et des jeux de données!

**Il y a de
nombreuses
organisations
d'enregistrement
et de types de
PID.**

ROR
GRID
ISNI

Établissements

ORCID (ISNI)
ScopusID
WoS
ResearcherID

Chercheurs

Crossref /DOI

Articles
Procédures
Monographies
*Jeux de données
Agences de financement
Subventions
Rapports
Normes
Préimprimés

Articles

Procédures

Monographies

*Jeux de données

Crossref /DOI

Agences de financement

Subventions

Rapports

Normes

Préimprimés

Datacite /DOI

Logiciels
Jeux de données
Collections
Documents audiovisuels
Évènements
Modèles

Datacite/DOI

Logiciels
Jeux de données
Collections
Documents audiovisuels
Évènements
Modèles

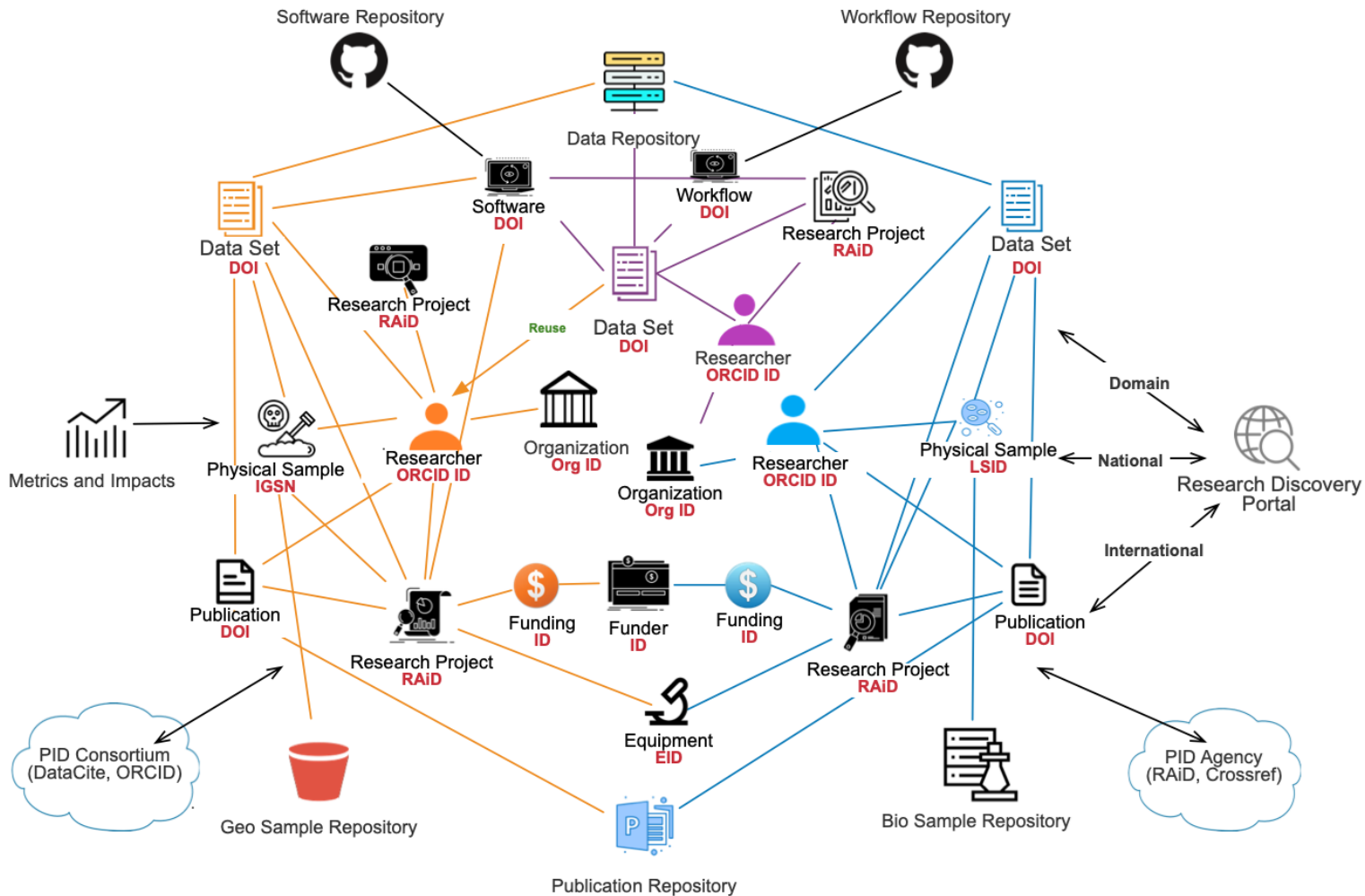
Raid

PID

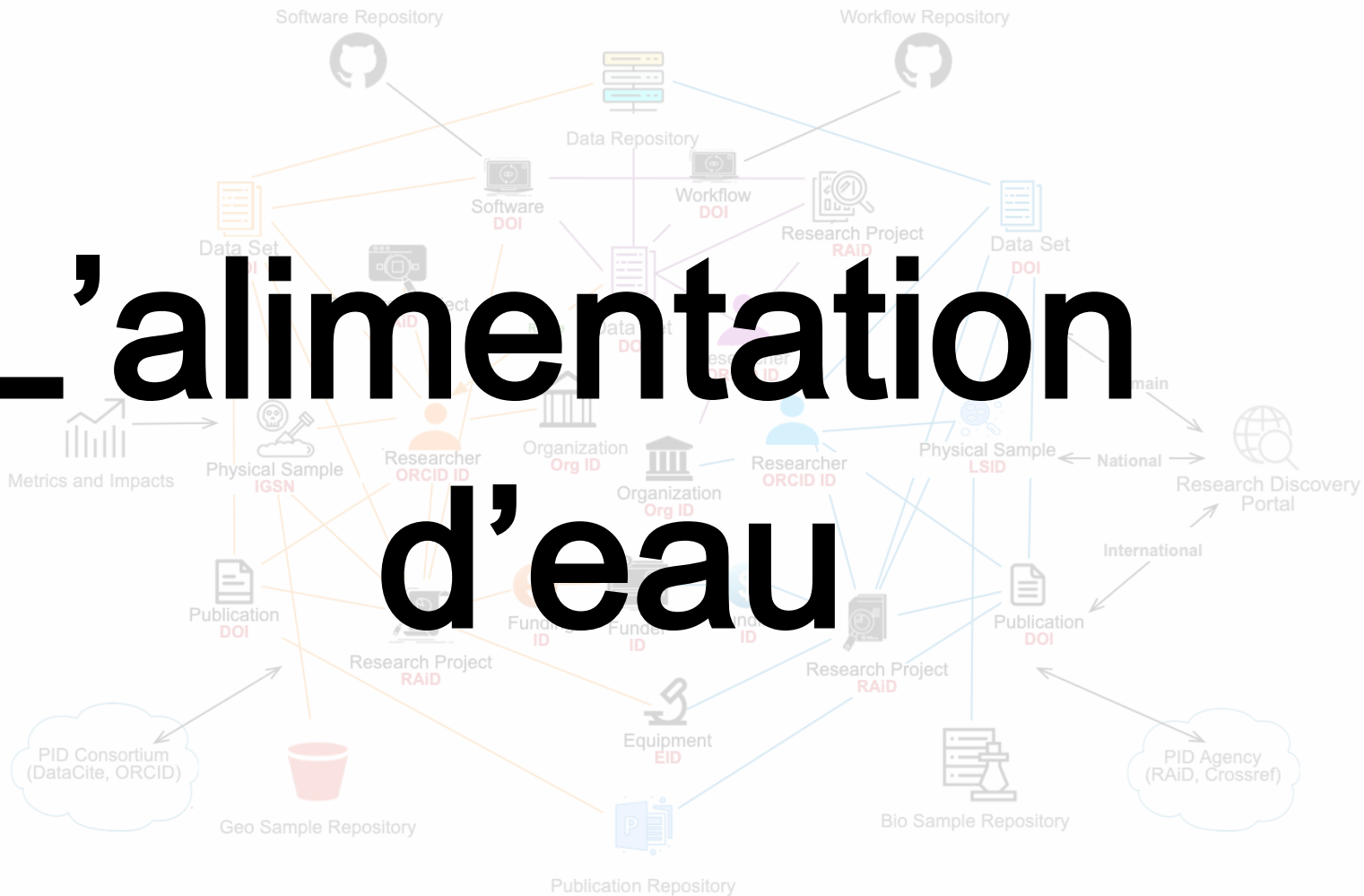
Toutes ces
plateformes
alimentent un
**pipeline ouvert de
métadonnées** avec
des données ou
extraient des
données de celui -ci.
(l'alimentation d'eau)

L'«API»

Interface de programmation
d'applications



L'alimentation d'eau



La plupart de ces organisations sont à but non lucratif (sauf Scopus et WoS évidemment).

**C'est bon?
Discutons de
quelques façons
dont l'eau circule!**

**Prenons un
exemple très
simple.**

Supposons que...

**Je configure mon
compte ORCID.**

**Je veux ajouter
mes publications!**

**Dans ORCID, je
peux vérifier les
API Crossref et
Datacite pour
trouver toutes les
publications
correspondant à
mon nom.**

**La plupart des
publications
attribuent des DOI.**

**Cette procédure
prendra du temps
la première fois et
ne fonctionnera
que si mes articles
ont des DOI.**

Mais...

Pour toutes les publications que je sais être les miennes et qui ont un DOI, les métadonnées sont automatiquement extraites de mon compte ORCID.

Et...!

**Maintenant que j'ai
un ORCID, ces
métadonnées sont
(idéalement)
incluses lorsque je
publie, ce qui
signifie que les
systèmes me
reconnaîtront.**

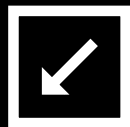
Métadonnées sur l'auteur

(non formatées)



Métadonnées sur l'éditeur

(jats ou comparables)



crossref (schéma crossref xml)



orcid (schéma orcid ... quasi Dublin, «bibtesque »)

Chaque schéma diffère un peu!

**Mike, je connais le
fonctionnement
d'ORCID.**

**Bien, faisons
comme si ...**

**J'ai demandé une
subvention à une
agence qui
possède un
compte ou une
intégration
ORCID.**

**Cette agence peut
alimenter mon
compte ORCID
avec de nouvelles
données.**

**Identifiants de financement
Identifiants de subvention
Jeux de données
Articles**

Et idéalement...

Quand je ferai une nouvelle demande de subvention, il me suffira de donner mon ORCID à l'agence pour qu'elle puisse extraire mes travaux sans que je doive remplir à nouveau le formulaire.

Si quelqu'un des
trois organismes
voit cette
présentation,
sachez que c'est
tout ce que je
veux que vous
reteniez. C'est la
seule chose .

Quand je ferai une
nouvelle demande de
subvention, je
donnerai simplement
mon ORCID à l'agence
pour qu'elle puisse
extraire mes travaux
sans que je doive
remplir à nouveau le
formulaire.

Prenons un *exemple*
plus complexe.

Cette fois, c'est au
tour des
«*bibliothèques*» *de*
jouer.

Supposons que...

**Mon établissement
utilise Unsub pour
savoir où ma faculté
publie et comment
cette pratique
correspond à nos
collections.**

Unsub a été créé et est entretenu par *deux personnes* seulement.

Le logiciel extrait les informations d'affiliation du Microsoft Academic Graph qui collecte les publications et utilise la correspondance des patrons issus du TLN.

**L'infrastructure
ouverte fait le
gros du travail ...**

**Il prend ensuite ces
données d'affiliation et
les compare avec l'API
Crossref pour l'ISSN et
les publications, la
collection que vous
avez fournie, les
identifiants
institutionnels GRID ou
ROR.**

**Unsub utilise
ensuite ces
données ...**

**... pour vous dire à quel
endroit vos chercheurs
publient, s'il s'agit de
publications en libre
accès (vérifie par rapport
à DOAJ et recherche les
politiques) et si les
revues auxquelles vous
êtes abonné ont des
publications .**

**Sans l'API
Crossref , tout ce
processus se
volatilise.**

**Les publications
qui n'utilisent pas
de DOI sont
essentiellement
«hors réseau».**

**Les publications
qui n'utilisent pas
de DOI sont
essentiellement
«hors réseau».**

**Par conséquent, beaucoup de gens
entrent les mêmes métadonnées
dans les systèmes manuellement. Ou
bien ils engagent des étudiants de
cycle supérieur pour le faire à leur
place. C'est une excellente utilisation
du temps de chacun, sans doute.**

**Les identifiants
pérennes nous
permettent d'avoir
une vue d'ensemble à
travers toutes ces
connexions et
interactions.**

Lorsque nous
parlons de soutien
aux PID, nous
parlons de soutien à
l'infrastructure
ouverte et à
l'échange libre de
métadonnées .

**Qu'en est -il de
tous les autres
objets?**

**Exact! Oui. Il y a trois (3)
règles générales.**

**Presque tous les endroits
principaux où les
chercheurs consignent
leurs travaux de nos jours
incorporent des PID quasi
automatiquement.**

Il y a de fortes chances que vous n'ayez jamais à vous préoccuper des PID institutionnels ou de l'ajout de DOI aux préimprimés .

1.33

Vous avez probablement déjà un identifiant ROR, et Arxiv gère automatiquement les DOI

La plupart du temps dans
le milieu des bibliothèques,
**les PID seront produits
sans votre intervention** .

1.66

Si vous hébergez du contenu qui ne se trouve nulle part ailleurs, ou si ce contenu est *principalement hébergé* sur un service que vous gérez, vous devriez produire (et enregistrer) un DOI pour ce contenu!

**Vous connaissez ce genre
de choses! Elles sont très
courantes dans les dépôts .**

Littérature grise

Rapports

*** Documents de travail**

Thèses

Projets

Diapositives

Vous ne devez pas
produire de DOI pour des
éléments qui en ont déjà
ailleurs.

**Autrement, vous empirez
les choses.**

Arrêtez.

2.5

L'entité avec laquelle vous
enregistrez vos PID
importe .

**Examinons
Crossref/Datacite de
nouveau.**

3.25

Crossref /DOI

Articles
Procédures
Monographies
Jeux de données*
Agences de financement
Subventions
Rapports
Normes
Préimprimés

Datacite /DOI

Logiciels
Jeux de données
Collections
Documents audiovisuels
Évènements
Modèles

Puisque leurs schémas sont spécialement conçus pour représenter certains types de contenu, la fidélité des métadonnées peut souffrir de la traduction d'un système à l'autre.

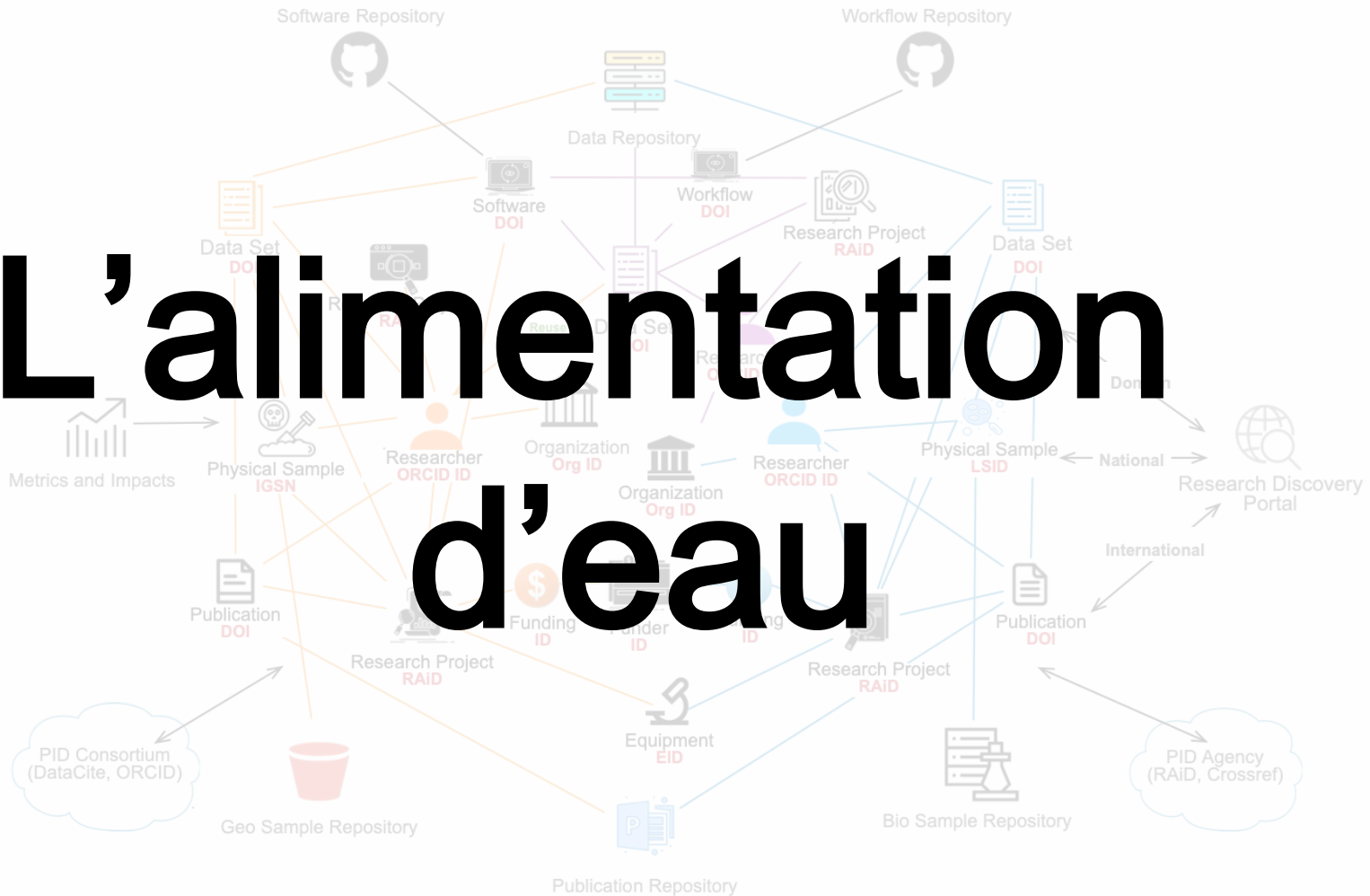
3.75

Crossref et Datacite sont des *amis* pour cette raison.

Médiocre à l'entrée.
Médiocre à la sortie .

3.81

L'alimentation d'eau



**«Je veux vraiment
être proactif
quant à
l'utilisation
judicieuse des
PID!»**

**– Vous, incroyablement,
maintenant.**

Plaider pour ORCID sans être trop insistant en rappelant aux gens que **c'est un peu comme un code -barres** .

Mettez l'accent sur le rôle du chercheur et la protection de la confidentialité.

Parfait, faites ceci :

Encouragez la culture des métadonnées là où vous le pouvez en aidant les chercheurs à comprendre pourquoi de bonnes métadonnées leur feront gagner du temps par la suite.

Vous allier d'abord à un bureau de recherche par **un argumentaire sur les mesures susceptibles d'être utilisées de manière abusive dans le cadre d'une évaluation scientifique** .

De grâce, **ne faites pas** ceci :

Ignorer les préoccupations des syndicats d'enseignants qui (à juste titre) ne sont pas ravis d'être réduits à des chiffres digestes.

Les **petit** s **caractère** s sont
si plaisants, n'est -ce pas?

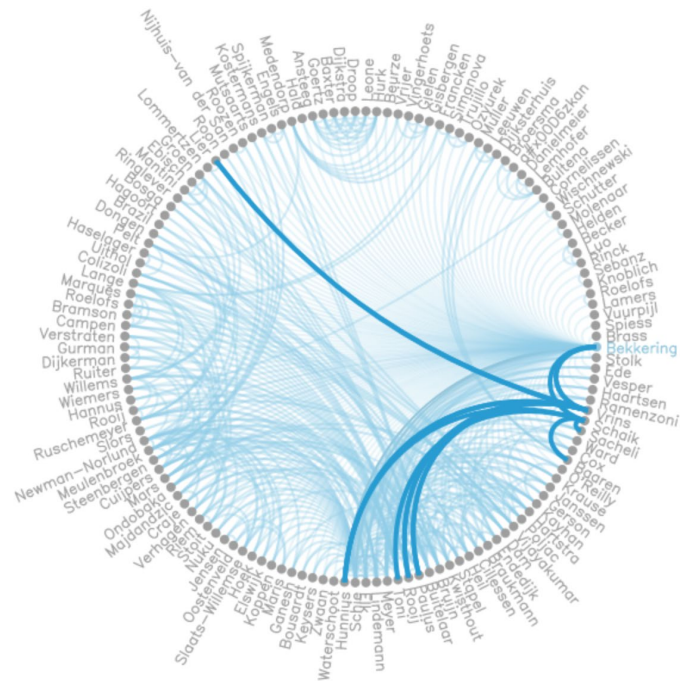
Je m'excuse de cette présentation
si rapide.

Je suis **Mike Nason** ,
Bibliothèque de l'UNB et le Public
Knowledge Project.

Je cède maintenant la parole à
Mark .

Rassembler les PID d'objets

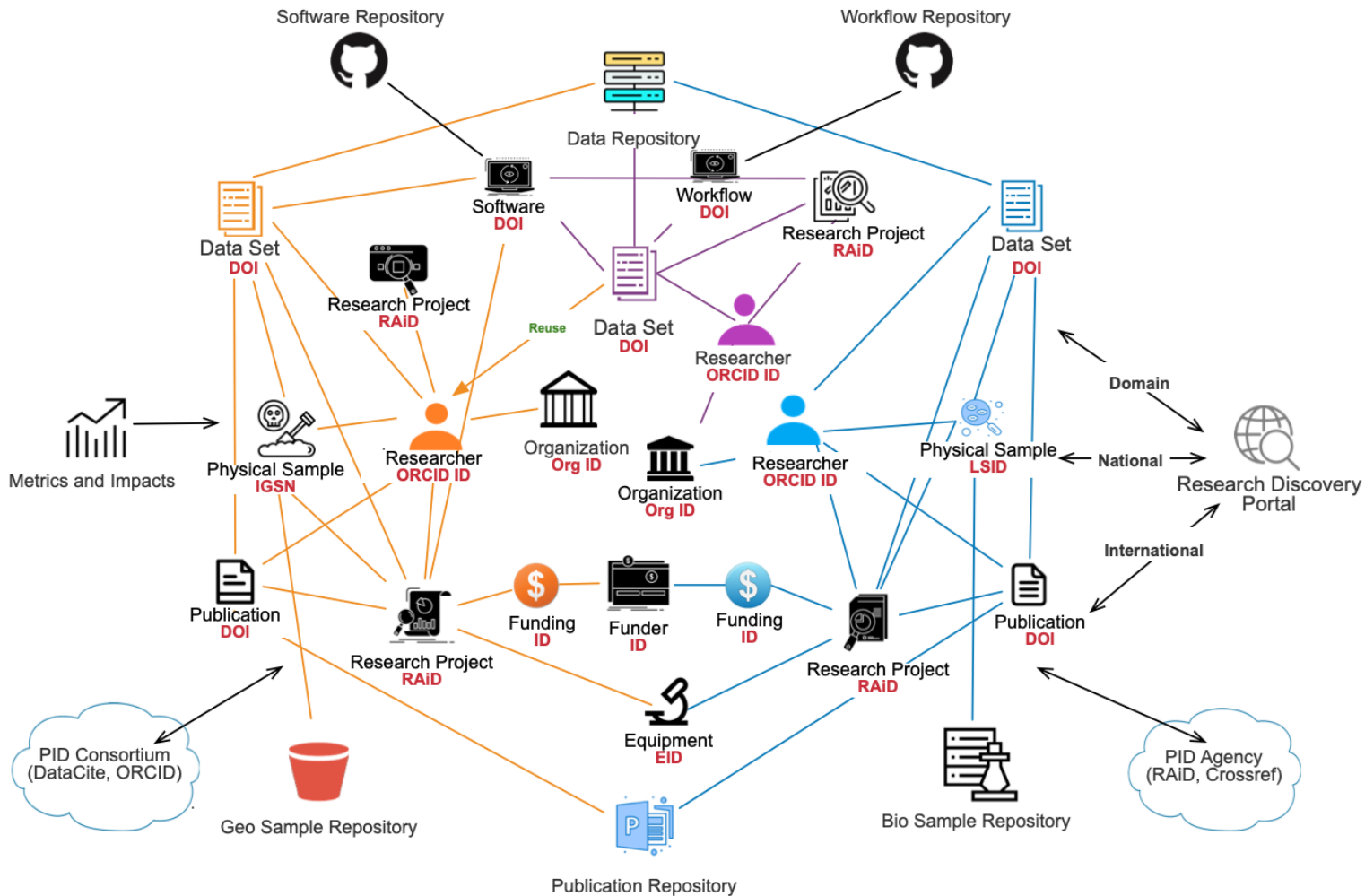
1. Les PID individuels sont utiles , mais...
2. Les PID sont surtout utiles dans un écosystème riche en PID, soit le graphique de PID
3. La création de liens entre tous les produits de recherche constitue la véritable valeur ajoutée.







<https://www.youtube.com/watch?v=yWOqeyPIVRo>



Défi 1 : Entente sur les PID

1. Il n'est pas possible de s'entendre sur **le PID qui les gouvernera tous** ...
2. Mais nous pouvons nous entendre sur les bonnes pratiques en matière de PID pour des objets particuliers.
3. Cette entente facilite la création et l'adoption de logiciels.
4. Elle permet ainsi la création d'un écosystème riche en PID facilitant la découverte des actifs dans l'écosystème de recherche.

Le PID en «chef»

1. Un identifiant d'activité de recherche (RaID) est un type unique de PID qui agit comme un agrégateur de divers PID, associé à un projet de recherche spécifique ou à une activité définie.
2. Le fait d'avoir un PID auquel on peut accéder de la même manière, via une API unique, fournit une représentation efficace et utile du graphique de PID.
3. On en voit pour les disciplines particulières, notamment le BioProject , mais les RaID ont le potentiel pour devenir les DOI de tous les projets de recherche.

Severe acute respiratory syndrome coronavirus 2

Accession: PRJNA686984 ID: 686984

SARS-CoV-2 Genome sequencing and assembly

NGS of PCR-tiled SARS-CoV-2

Accession	PRJNA686984
Data Type	Genome sequencing
Scope	Multisolate
Organism	Severe acute respiratory syndrome coronavirus 2 [Taxonomy ID: 2697049] Viruses; Riboviria; Orthornavirae; Pisuviricota; Pisoniviricetes; Nidovirales; Cornidovirineae; Coronaviridae; Orthocoronavirinae; Betacoronavirus; Sarbecovirus; Severe acute respiratory syndrome-related coronavirus; Severe acute respiratory syndrome coronavirus 2
Grants	"CK19-1904 Epidemiology and Laboratory Capacity for Infectious Diseases (ELC)" (Grant ID 6 NU50CK000552-02-01, Center for Disease Control and Prevention)
Submission	Registration date: 21-Dec-2020 Colorado Department of Public Health and Environment

Project Data:

Resource Name	Number of Links
SEQUENCE DATA	
Nucleotide (Genomic RNA)	1962
SRA Experiments	1971
Protein Sequences	23526
OTHER DATASETS	
BioSample	1971

▾ SRA Data Details

Parameter	Value
Data volume, Gbases	138
Data volume, Mbytes	72794

See [Genome Information for Severe acute respiratory syndrome-related coronavirus](#)

NAVIGATE UP

This project is a component of the [COVID-19 Outbreak](#)

This project is a component of the [INSDC SARS-CoV-2 Viral Sequencing Data](#)

NAVIGATE ACROSS

29 additional projects are components of the [COVID-19 Outbreak](#).

250 additional projects are components of the [INSDC SARS-CoV-2 Viral Sequencing Data](#).

134 additional projects are related by organism.

Related information[BioProject](#)[BioSample](#)[Genomic RNA](#)[Nucleotide](#)[Protein](#)[SRA](#)[Taxonomy](#)[Umbrella projects](#)**Recent activity**[Turn Off](#) [Clear](#)

 [Severe acute respiratory syndrome coronavirus 2](#) BioProject

 [615625\[top bioproject\] NOT 718231\[uid\] \(29\)](#) BioProject

 [Lichtheimia corymbifera strain:B63a](#) BioProject

 [Severe acute respiratory syndrome coronavirus 2 isolate Wuhan-Hu-1, c](#) Nucleotide

 [covid-19 \(71327\)](#) Nucleotide

[See more...](#)

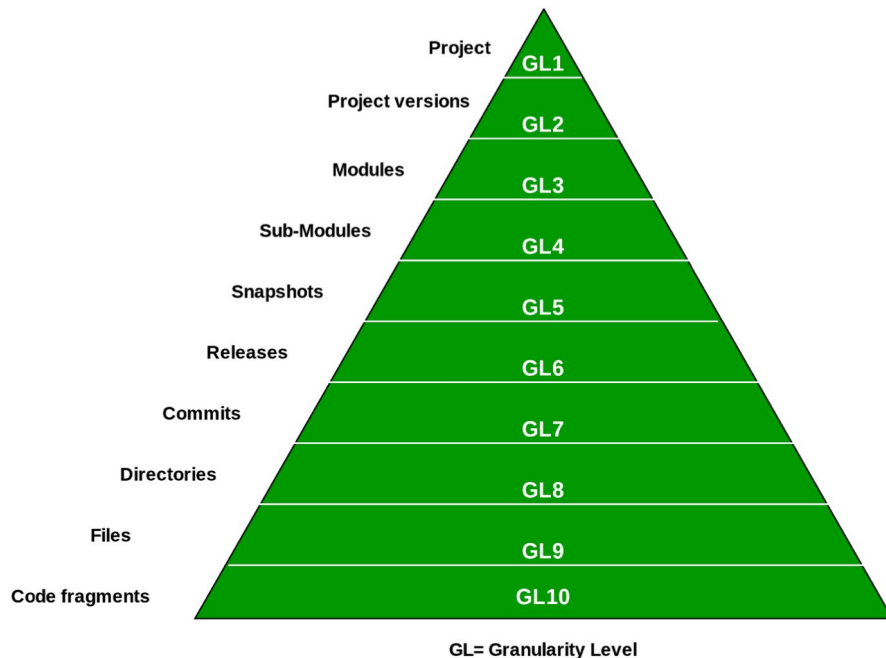
Identifiants d'activité de recherche (RaID)

1. Les RaID proviennent de travaux en Australie pour la création d'«enregistrements de gestion de recherche».
2. Ils sont des agrégateurs de PID pour toutes les ressources associées à un projet précis.
3. Ils sont en cours de révision et d'approbation de l'ISO (cette année).
4. Les RaID seront produits par des partenaires régionaux et offerts gratuitement.

Vidéo sur les RaID

Autres objets de PID : logiciels

1. Les PID de logiciels sont compliqués!
2. Intrinsèques
 - a. PID générés par l'environnement de développement des logiciels, tels que VCS (identifiants Git)
 - b. Bases des [SWHID](#)
3. Extrinsèques
 - a. PID en dehors du contexte des logiciels.
 - b. DOI créés par des dépôts tels que Zenodo



Autres objets de PID : matériel

1. «Pour interpréter les jeux de données numériques, il faut beaucoup d'informations sur le matériel utilisé pour générer les données, qu'il s'agisse de réseaux de capteurs ou de machines de laboratoire».
2. Le groupe de travail RDA PIDINST a créé un schéma avec 43 éléments comprenant les identifiants, les propriétaires, les fabricants, les variables mesurées, etc.
3. 2 exemples
 - a. DOI DataCite
 - b. Cadre ePIC
4. Les identifiants RRID sont aussi utilisés pour le matériel et les installations

Autres objets de PID : ressources

1. RRID : identifiants de ressources de recherche
 - a. lignées cellulaires, anticorps, plasmides, organismes modèles, installations et matériel
2. Rigueur/détail supplémentaire dans la description des ressources de référence associées, typiquement dans le *Matériel*
3. Utilisation grandissante par les éditeurs de revues



Bien d'autres dans des contextes de domaines

MycoBank ID

InChi

DIN

EC Number

PDP ID

Avenir?

1. À court terme, il faudrait une adoption généralisée des PID fondamentaux les plus courants, y compris les RaID
2. Identifiants décentralisés W3Cs ([DIDs](#))
 - a. *«Un nouveau type d'identifiant permettant une identité numérique vérifiable et décentralisée. Les DID identifient tous les objets (par exemple, une personne, une organisation, une chose, un modèle de données, une entité abstraite, etc.) que le contrôleur des DID décide d'identifier».*
3. Jetons non fongibles (NFT)
 - a. Ont-ils un rôle dans la communication scientifique et l'identité pérenne? (consultez [Scholarly Kitchen](#))