

# demystifying the automagical

an honest appraisal of the challenges in persistent identifiers and  
open scholarly infrastructure



# démystifier l'automatisme

une évaluation honnête des défis posés par les identificateurs  
persistants et l'infrastructure savante ouverte

introductions

---

~

I'm the **open scholarship & publishing librarian** (aka, I guess, "scholarly communications") at the University of New Brunswick.

Je suis **bibliothécaire chargé des bourses et de l'édition** (alias, je suppose, « communications savantes ») à l'université du Nouveau-Brunswick.

I also work for **PKP** as a member of their publishing services team, where I am the **Crossref/metadata liaison**.

I stand at a pretty interesting professional intersection between publishing, authoring, and discovery.

That means I more or less **never shut up about open scholarly infrastructure!**

Je travaille également pour **PKP** en tant que membre de l'équipe des services d'édition, où j'assure la liaison entre **Crossref et les métadonnées**.

Je me trouve à une intersection professionnelle très intéressante entre l'édition, la rédaction et la découverte.

Cela signifie que je **ne me tais pratiquement jamais au sujet de l'infrastructure savante ouverte !**

Comme beaucoup de gens qui ne se taisent pas, je ressemble à ça :

I am a [white, cis] settler from the unceded (aka, stolen) territory of the **Mi'kmaq-Wolastoquey peoples** just a short hop from the Wolastoq River. Les colons de la région ont rebaptisé cette rivière « Saint John River », ce qui témoigne à la fois de leur *répression et de leur manque de créativité*.



demystifying

---

démystifier

## some resources

We don't really have time today to go super deep on how PIDs work in practice. But, if you're curious about this in your own time, I've been around the block.

Getting Found, Staying Found.  
Persistent Identifiers & Their Value.  
([deck](#) | [video](#))

FAIR Principles & Persistent Identifiers  
([deck](#) | [video](#))

## quelques ressources

Nous n'avons pas vraiment le temps aujourd'hui d'approfondir le fonctionnement pratique des PID. Mais si vous êtes curieux de savoir ce qu'il en est, sachez que j'ai fait le tour de la question.

Se faire repérer, rester repéré. Les identifiants persistants et leur valeur.  
([présentation](#) | [vidéo](#))

Les principes FAIR et les identificateurs persistants ([présentation](#) | [vidéo](#))

# what is a PID?

PID is short for *Persistent Identifier*. Half of this should make complete sense to you. We have identifiers for *all kinds of things, all the time*.

- social insurance numbers
- driver's licenses
- license plates
- medicare
- student ids

# qu'est-ce que le PID ?

PID est l'abréviation de Persistent Identifier (identifiant persistant). La moitié de ce qui précède devrait vous sembler tout à fait logique. Nous avons des identifiants pour toutes sortes de choses, tout le temps.

- les numéros d'assurance sociale
- les permis de conduire
- les plaques d'immatriculation
- l'assurance-maladie
- cartes d'étudiant



# what is a PID?

Identifiers typically refer to physical objects and are often created/managed locally. They're useful for record-keeping and data retrieval/searchability. They're useful for disambiguation, too!

For example, there is more than one Mike Nason in Canada, but only *one* of them has my social insurance number.

# qu'est-ce que le PID ?

Les identificateurs se réfèrent généralement à des objets physiques et sont souvent créés/gérés localement. Ils sont utiles pour l'archivage et la recherche de données. Ils sont également utiles pour la désambiguïsation !

Par exemple, il y a plus d'un Mike Nason au Canada, mais un seul d'entre eux a mon numéro d'assurance sociale.

# persistent

Persistence (not to be confused with "permanence", please) means "long-lasting"... a function of the life of the service that has stewardship over a kind of identifier.

# persistante

La persistance (à ne pas confondre avec la « permanence », s'il vous plaît) signifie « longue durée »... une fonction de la vie du service qui a la responsabilité d'un type d'identifiant.

# identifier

A unique string or pattern referring to an object, person, document, file, website, skyscraper, sunflower seed, yawn, feature film, bicycle, unicorn... whatever.

*You get it.*

# identificateur

Chaîne ou motif unique faisant référence à un objet, une personne, un document, un fichier, un site web, un gratte-ciel, une graine de tournesol, un bâillement, un long métrage, une bicyclette, une licorne... peu importe.

Vous l'aurez compris.

In Librarianship, we use *a sort of dizzying array* of persistent identifiers.

---

En bibliothéconomie, nous utilisons une sorte d'éventail vertigineux d'identifiants permanents.

- issn
- isbn
- doi
- handle
- ark
- orcid
- ror
- scopusid
- ringgold

- magnet link
- urn
- uri
- xri
- purl
- viaf
- isni
- oclc #
- and so on...

And, **in the context of publishing and open scholarly infrastructure**, we can narrow this down into some especially vital PIDs.

---

Et, **dans le contexte de l'édition et de l'infrastructure universitaire ouverte**, nous pouvons réduire cette liste à quelques PID particulièrement importants.

- doi
- orcid
- ror

This is a little reductive,  
and we'll get to that.

---

C'est un peu réducteur,  
et nous y reviendrons.

DOIs are **the big one**, so let's pivot to those real quick.

---

Les DOI sont les plus **importants**, et nous allons donc les aborder rapidement.



**Oh! But!** Just because DOIs are an industry standard *for academic publications* doesn't mean that's specifically what they're for. *You can assign a DOI to basically anything.*

---

Mais ce n'est pas parce que les DOI sont une norme industrielle *pour les publications universitaires* qu'ils ne servent qu'à cela. *Vous pouvez attribuer un DOI à pratiquement n'importe quoi.*

# DOIs

DOIs are ubiquitous. We see them *all over the place*:

- in references/bibliographies
- on article/journal websites
- in repositories
- on published datasets

And, we *probably* know one handy thing about them: ***If you click on a DOI that looks like a link, it will take you to the thing.***

Les DOI sont omniprésents. On les voit partout :

- dans les références/bibliographies
- sur les sites web des articles/journaux
- dans les référentiels
- sur les ensembles de données publiés

Nous connaissons probablement une chose pratique à leur sujet : ***Si vous cliquez sur un DOI qui ressemble à un lien, il vous conduira à l'objet en question.***

DOIs are the **most prominent** persistent identifier.

They are also, arguably, the **most important** persistent identifier.

Les DOI sont l'identifiant permanent le **plus connu**.

Ils sont aussi, sans doute, l'identifiant permanent le **plus important**.

# how DOIs work / fonctionnement des DOI

A DOI is made up of two chunks, et ils ont des significations différentes.

**prefix / suffix**  
**10.4324 / 9780203051238-5**

A **prefix** is usually associated with a publisher or organization. DOIs for that organization will usually have the same prefix. A **suffix** is meant to be a machine-readable (*not* human-readable), opaque, unique string that is specific to the singular work to which it is assigned.

**prefix / suffix**  
**10.4324 / 9780203051238-5**

Un **préfixe** est généralement associé à un éditeur ou à une organisation. Les DOIs de cette organisation auront généralement le même préfixe. Un **suffixe** est une chaîne de caractères unique, opaque, lisible par une machine (et non par un être humain) et spécifique à l'œuvre singulière à laquelle il est attribué.

# how DOIs work / fonctionnement des DOI

Prepending DOI with <https://doi.org/>, it turns into a URL.

<https://doi.org/10.4324/9780203051238-5>

[doi.org/10.4324/9780203051238-5](https://doi.org/10.4324/9780203051238-5)

Clicking this will redirect me to the publication this DOI is associated with.

The process of a DOI redirecting you to a publication is called **resolution**.

And resolution is facilitated by **registration agencies**.

[doi.org/10.4324/9780203051238-5](https://doi.org/10.4324/9780203051238-5)

En cliquant sur ce lien, je serai redirigé vers la publication à laquelle ce DOI est associé.

Le processus par lequel un DOI vous redirige vers une publication s'appelle la **résolution**.

La résolution est facilitée par les **agences d'enregistrement**.

# how DOIs work / fonctionnement des DOI

Les DOI *ne sont pas des liens bit.ly ou tiny.url.*

Basic redirect/link-shortening services like bit.ly are for one purpose only, really. They provide a shorter link for more convenient use/readability, and they then direct a user to the original url.

For example, here's a bit.ly link for this talk in the program:

<https://bit.ly/3yvEzAZ>

Les services de redirection/raccourcissement de lien de base tels que bit.ly n'ont qu'un seul but, en réalité. Ils fournissent un lien plus court pour une utilisation/lisibilité plus pratique, et dirigent ensuite l'utilisateur vers l'url d'origine.

Par exemple, voici un lien bit.ly pour cet exposé dans le programme :

<https://bit.ly/3yvEzAZ>

# how DOIs work / fonctionnement des DOI

Si vous cliquez sur un DOI qui ressemble à un lien, vous serez redirigé vers ce dernier.

But, plenty of people treat DOIs this way. Or, they assume it's their only real function. **This is complicated by the fact that some PIDs *do*, essentially, work this way.**

Or, swinging wildly in the other direction, they ascribe *fresh meaning* to DOIs and make some wild presumptions about DOIs as a sign of scholarly legitimacy.

De nombreuses personnes traitent les DOI de cette manière, ou supposent que c'est leur seule fonction réelle. **Cette situation est compliquée par le fait que les « handles » fonctionnent essentiellement de cette manière.**

Ou bien, dans l'autre sens, ils attribuent une nouvelle signification aux DOI et font des suppositions farfelues sur les DOI en tant que signe de légitimité scientifique.

A DOI is *a lot* more than a  
basic redirect.

---

Un DOI est *bien plus* qu'une  
simple redirection.



Any single DOI is a reference to **an entire publication record**. That publication record is **full of metadata**.

And, *one* of these metadata elements is the **publication's URL**.

When you **resolve a doi** by clicking on it:

- the record is accessed
- the stored URL is retrieved
- you are sent to the stored URL

The URL can be updated by the publisher. The DOI is ***persistent*** even though the URL changes.

Chaque DOI est une référence à **un enregistrement de publication entier**. Cette fiche de publication est **pleine de métadonnées**.

L'un de ces éléments de métadonnées est **l'URL de la publication**.

Lorsque vous résolvez un DOI en cliquant dessus :

l'enregistrement est consulté  
l'URL stockée est récupérée  
vous êtes renvoyé à l'URL stockée

L'éditeur peut mettre à jour l'URL. Le DOI reste inchangé. Le DOI est ***persistant*** même si l'URL change.

Depending on the registration agency that DOI is registered to, and what kind of work it is, there's a boatload of other metadata you can access.

---

Selon l'agence d'enregistrement à laquelle le DOI est enregistré et le type de travail, il existe une multitude d'autres métadonnées auxquelles il est possible d'accéder.

Unlike publications themselves, metadata is *typically* free. And, we can learn a lot from it. Crossref, for example, can store the following things (not inclusive) as **publicly accessible metadata**.

---

Contrairement aux publications elles-mêmes, les métadonnées sont généralement gratuites. Nous pouvons en tirer beaucoup d'enseignements. Crossref, par exemple, peut stocker les éléments suivants (non exhaustifs) en tant que **métadonnées accessibles au public**.

title	titre
subtitle	sous-titre
<i>authors</i>	<i>auteurs</i>
<i>orcid</i>	<i>orcid</i>
affiliation	affiliation
<i>copyright license</i>	<i>licence de copyright</i>
funder/grant ids	identifiant du bailleur de fonds
languages	langues
<i>ror</i>	<i>ror</i>
<i>references</i>	<i>références</i>
<i>resource location</i>	<i>emplacement de la ressource</i>
version	version
publisher	éditeur
journal/volume/issue	journal/volume/numéro
<i>related doi</i>	<i>dois apparenté</i>
dates	dates
abstracts	résumés



```
1 <crossref_result version="3.0" xsi:schemaLocation="http://www.crossref.org/qrschema/3.0
  http://www.crossref.org/schemas/crossref_query_output3.0.xsd">
2 <query_result>
3 <head>
4 <doi_batch_id>none</doi_batch_id>
5 </head>
6 <body>
7 <query status="resolved">
8 <doi type="journal_article">10.1371/journal.pone.0297222</doi>
9 <crm-item name="publisher-name" type="string">Public Library of Science (PLOS)</crm-item>
10 <crm-item name="prefix-name" type="string">Public Library of Science</crm-item>
11 <crm-item name="member-id" type="number">340</crm-item>
12 <crm-item name="citation-id" type="number">159115887</crm-item>
13 <crm-item name="journal-id" type="number">60407</crm-item>
14 <crm-item name="deposit-timestamp" type="number">202403211021060000</crm-item>
15 <crm-item name="owner-prefix" type="string">10.1371</crm-item>
16 <crm-item name="last-update" type="date">2024-03-21T17:21:43Z</crm-item>
17 <crm-item name="created" type="date">2024-03-21T17:21:14Z</crm-item>
18 <crm-item name="citedby-count" type="number">0</crm-item>
19 <doi_record>
20 <crossref xsi:schemaLocation="http://www.crossref.org/xschema/1.1 http://doi.crossref.org/schemas/unixref1.1.xsd">
21 <journal>
22 <journal_metadata language="en">
23 <full_title>PLOS ONE</full_title>
24 <abbrev_title>PLOS ONE</abbrev_title>
25 <issn media_type="electronic">1932-6203</issn>
26 </journal_metadata>
27 <journal_issue>
28 <publication_date media_type="online">
29 <month>3</month>
30 <day>21</day>
31 <year>2024</year>
32 </publication_date>
33 <journal_volume>
34 <volume>19</volume>
35 </journal_volume>
```

This metadata is, as you might  
have guessed, hugely useful!

---

Ces métadonnées sont, comme vous  
l'avez deviné, extrêmement utiles !

# dans Crossref

- article-level record
- all titles on a given prefix
- all works registered to an ORCID
- works affiliated with grants
- most-cited by prefix

These are all calls against the Crossref public API for metadata related to publications. This is just a tiny fraction of the sorts of queries you could make.

- enregistrement au niveau de l'article
- tous les titres d'un préfixe donné
- toutes les œuvres enregistrées auprès d'un ORCID
- ouvrages affiliés à des subventions
- les plus cités par préfixe

Il s'agit d'appels à l'API publique de Crossref pour les métadonnées liées aux publications. Il ne s'agit que d'une infime partie des types de requêtes que vous pouvez effectuer.

Oooh, that's a lot of data! Imagine  
what you could do with it!

---

Oooh, cela fait beaucoup de données !  
Imaginez ce que vous pourriez en faire !

You know... besides the persistent  
access, disambiguation, and  
discoverability stuff.

---

Vous savez... en dehors de l'accès  
permanent, de la désambiguïsation et  
de la découvrabilité.



## when connected...

- I could have publications automatically added to my **ORCID** record.
- I could find all the works in **OpenAire** written by researchers at my institution.
- I could get publishing metrics about the journals my faculty publish in via **OpenAlex**.

## lorsqu'il est connecté...

- Je pourrais faire ajouter automatiquement des publications à mon dossier **ORCID**.
- Je pourrais trouver dans **OpenAire** tous les travaux écrits par des chercheurs de mon institution.
- Je pourrais obtenir des données sur les revues dans lesquelles mes professeurs publient par l'intermédiaire d'**OpenAlex**.

## when connected...

- I could **reveal relationships** between a funder, a specific grant ID, and all the various products of scholarship created/disseminated as a result.
- I could **evaluate the "completeness" of my published metadata** with a variety of tools.

## lorsqu'il est connecté...

- Je pourrais **révéler les relations** entre un bailleur de fonds, un numéro de subvention spécifique et tous les divers produits d'érudition créés/diffusés en conséquence.
- Je pourrais **évaluer l'« exhaustivité » de mes métadonnées publiées** à l'aide de divers outils.

We'll come back to this.



Nous y reviendrons.

# agences d'enregistrement

Persistent identifiers are managed by **registration agencies** (typically international not-for-profits) that store records/metadata, facilitate resolution requests, and may or may not offer other services based on membership.

They do much of this through APIs.

Les identificateurs persistants sont gérés par des **agences d'enregistrement** (généralement des organisations internationales à but non lucratif) qui stockent les enregistrements/métadonnées, facilitent les demandes de résolution et peuvent ou non offrir d'autres services en fonction de l'appartenance à l'organisation.

La plupart de ces services sont assurés par des interfaces de programmation (API).

# agences d'enregistrement

There are **a lot of registration agencies!**

Each agency may differ in mandate, governance, scope, service, supported objects, membership terms, and feature set.

They also, often, work together and share data.

Il existe **de nombreuses agences d'enregistrement !**

Chaque agence peut différer en termes de mandat, de gouvernance, de portée, de service, d'objets pris en charge, de conditions d'adhésion et d'ensemble de fonctionnalités.

Elles travaillent aussi souvent ensemble et partagent des données.

# PIDs for scholarly works

Publishing-flavoured / Au goût de l'édition

## Crossref (DOI)

La plupart des éditeurs universitaires sont membres de Crossref. Au moment où j'écris ces lignes (21 mai), Crossref a enregistré **158 544 829** DOI.

Crossref are a *big deal*.

articles	articles
proceedings	actes
monographs	monographies
*datasets	*fichiers de données
funding agencies	agences de financement
grants	subventions
reports	rapports
standards	normes
preprints	préimpressions

# PIDs for scholarly works

Repositories-flavoured / Les référentiels à saveur

## Datacite (DOI)

Bien que certains éditeurs universitaires utilisent Datacite pour les DOI des articles, il est beaucoup plus couramment utilisé dans les dépôts de données/institutionnels/disciplinaires. Datacite et Crossref travaillent ensemble pour relier les données de recherche aux publications.

software	logiciel
datasets	ensembles de données
collections	collections
audio/visual	audio/visuel
event	événement
model	modèle
*publications	*publications

# PIDs for researchers

ORCID  
Scopus ID  
WoS Researcher ID

ORCID are the go-to here, with Scopus and WoS offerings both restricted to publications present on those platforms. However, these services can share data between them.

chercheurs

ORCID est la référence en la matière, les offres de Scopus et WoS étant toutes deux limitées aux publications présentes sur ces plateformes. Toutefois, ces services peuvent partager des données entre eux.



# PIDs for organizations

ROR  
GRID  
Ringgold

organisations

The predominant use-case for organizational IDs is in strengthening connections between records using open scholarly infrastructure.

Le principal cas d'utilisation des identifiants organisationnels est le renforcement des connexions entre les documents à l'aide d'une infrastructure savante ouverte.

# agences d'enregistrement

Registration agencies **provide metadata schema** through which users can describe the objects they are registering PIDs for.

As you can imagine, **you'd describe a person differently than you'd describe a dataset, or a journal article, or an organization.** Even when agencies use the same type of PID (like the DOI), the schema they use may vary.

Les agences d'enregistrement **fournissent des schémas** de métadonnées permettant aux utilisateurs de décrire les objets pour lesquels ils enregistrent des PID.

Comme vous pouvez l'imaginer, **vous décrirez une personne différemment d'un ensemble de données, d'un article de revue ou d'une organisation.** Même lorsque les agences utilisent le même type de PID (comme le DOI), le schéma qu'elles utilisent peut varier.

Obviously, metadata is hugely important to the utility of PIDs!

---

Il est évident que les métadonnées sont extrêmement importantes pour l'utilité des PID !

The metadata we get out of these systems – and its *utility* – is **very much dependent on its quality**.

We have a general expression in the metadata universe. That's "**garbage in, garbage out**."

Metadata is kind of **everyone's responsibility**. Researchers, librarians, publishers, registration agencies... everyone has a stake in accurate, usable metadata.

Les métadonnées que nous tirons de ces systèmes - et leur utilité - **dépendent fortement de leur qualité**.

Il existe une expression générale dans l'univers des métadonnées. Il s'agit de l'expression « **garbage in, garbage out** ».

Les métadonnées sont en quelque sorte **la responsabilité de tous**. Les chercheurs, les bibliothécaires, les éditeurs, les agences d'enregistrement... tout le monde a intérêt à ce que les métadonnées soient précises et utilisables.

## PIDs aren't

- just links to things
- academic bit.ly
- status symbols
- permanent\*\*\*
- magic (at least not literally)

## PIDs are

- potentially huge time savers
- useful for finding research
- interconnected between services
- the backbone of open scholarly infrastructure

## les PID ne sont pas

- juste des liens vers des choses
- bit.ly académique
- symboles de statut
- permanent\*\*\*
- magie (du moins pas littéralement)

## les PID sont

- un gain de temps potentiellement
- utiles pour la recherche
- interconnectés entre les services
- l'épine dorsale de l'infrastructure savante ouverte

Congratulations, **you now know *more* about DOIs** than a frankly surprising amount of people.

And, by *extension*, **you now know more about PIDs** than a frankly surprising amount of people.

Félicitations, **vous en savez maintenant plus sur les DOIs** qu'un nombre franchement surprenant de personnes.

Et, par extension, **vous en savez maintenant plus sur les PIDs** qu'un nombre franchement surprenant de personnes.

Let's discuss the "automagical".  
Discutons de « l'automagique ».

# "automagical"

## automatic + magical

"Automagical" is a colloquial term that describes things that happen "automatically", and, in a way that's either obfuscated, inscrutable, or not well understood by the user. As a result, the results appear to be "magic".

## automatique + magique

« Automagical » est un terme familier qui décrit des choses qui se produisent « automatiquement », et d'une manière qui est soit obscurcie, soit impénétrable, soit mal comprise par l'utilisateur. En conséquence, les résultats semblent être « magiques ».



# "automagical"

Most of what we think of as "automagical" in this space is likely the result of one of the following:

1. **Unseen/misunderstood labour** | Sometimes a *person* is manually depositing PID metadata they've typed out by hand, record by record.
2. **An API** | Or, a system does it automatically.

La plupart de ce que nous considérons comme « automatique » dans cet espace est probablement le résultat de l'un des éléments suivants :

1. **Travail invisible/méconnu** | Parfois, une personne dépose manuellement les métadonnées PID qu'elle a saisies à la main, enregistrement par enregistrement.
2. **Une API** | Ou un système le fait automatiquement.

# APIs

Application  
Programming  
Interface

An API is, basically, a set of rules for interacting with software.

Think of it as being a little like a translator working as an intermediary between two people who don't speak the same language.

Application  
Programming  
Interface

Une API est, en fait, un ensemble de règles permettant d'interagir avec un logiciel.

C'est un peu comme un traducteur qui sert d'intermédiaire entre deux personnes qui ne parlent pas la même langue.

**APIs are everywhere.** When my calendar app tells me today's forecast, it's accessing that information using the Accuweather API. When my watch vibrates because I got a text message, that's because Garmin's API is communicating with Apple's notifications API.

APIs are how disparate systems, built by different people, using different languages and definitions, find common ground and share information.

**Les API sont partout.** Lorsque mon application de calendrier m'indique les prévisions pour aujourd'hui, elle accède à ces informations à l'aide de l'API Accuweather. Lorsque ma montre vibre parce que j'ai reçu un SMS, c'est parce que l'API de Garmin communique avec l'API de notification d'Apple.

Les API permettent à des systèmes disparates, construits par des personnes différentes, utilisant des langages et des définitions différents, de trouver un terrain d'entente et de partager des informations.

# open scholarly infrastructure

This network of APIs connecting these PID registration agencies to other, open services is, increasingly, infrastructure relied upon by researchers and institutions, whether they are really aware of it.

**Almost all open scholarly infrastructure is based around APIs and the ability to easily push and pull this metadata around.**

Ce réseau d'API reliant ces agences d'enregistrement de PID à d'autres services ouverts est, de plus en plus, une infrastructure sur laquelle comptent les chercheurs et les institutions, qu'ils en soient réellement conscients ou non.

**Presque toute l'infrastructure universitaire ouverte est basée sur des API et sur la possibilité de pousser et de tirer facilement ces métadonnées.**



Let's do an example.



Prenons un exemple.



peer-review





peer-review



self-archiving





Publishing is more iterative (and tied to a wider range of accessible outputs) than it's ever been.

- users might post their work on a preprint server before peer review.
- funder mandate might require open access or, at least, an open access *version* of the work.
- data management might require publication of a data set.
- these things are true for *every* co-author.

La publication est plus itérative (et liée à un plus large éventail de résultats accessibles) qu'elle ne l'a jamais été.

- Les utilisateurs peuvent publier leurs travaux sur un serveur de préimpression avant l'examen par les pairs.
- le mandat du bailleur de fonds peut exiger le libre accès ou, au moins, une version du travail en libre accès
- la gestion des données peut exiger la publication d'un ensemble de données.
- ces éléments s'appliquent à tous les coauteurs.



You can see how even something  
as benign as a typo might end up  
in a lot of places.

---

Vous pouvez voir comment même  
quelque chose d'aussi bénin qu'une  
faute de frappe peut se retrouver à de  
nombreux endroits.

We are effectively shotgunning  
metadata across the internet.  
Often automatically.

---

En fait, nous vomissons des  
métadonnées sur l'internet.  
Souvent de manière automatique.

There are *so many* "points of publication" that consistency is a challenge.

---

Il y a tellement de « points de publication » que la cohérence est un défi.

It's almost like metadata is *a little bit out of our hands*. Not *totally*. But, like, a little bit!

---

On a l'impression que les métadonnées nous échappent un peu. Pas totalement. Mais un peu !





### Traffic Manager

#### Water

Water Availability



Sewage Treatment



Water consumption: 1,498,112 m<sup>3</sup>/week  
Water pumping capacity: 2,340,000 m<sup>3</sup>/week

Sewage production: 1,497,152 m<sup>3</sup>/week  
Sewage draining capacity: 2,240,000 m<sup>3</sup>/week

#### Legend

- Water Facilities: ■ Active ■ Inactive
- Sewage Facilities: ■ Active ■ Inactive
- Buildings:
  - Water & Sewage
  - Water only & No sewage
  - No water & no sewage

Briarwood District

Garnet Heights

Chester District

Washington Park

Construction cost: 2500

Rockvalley

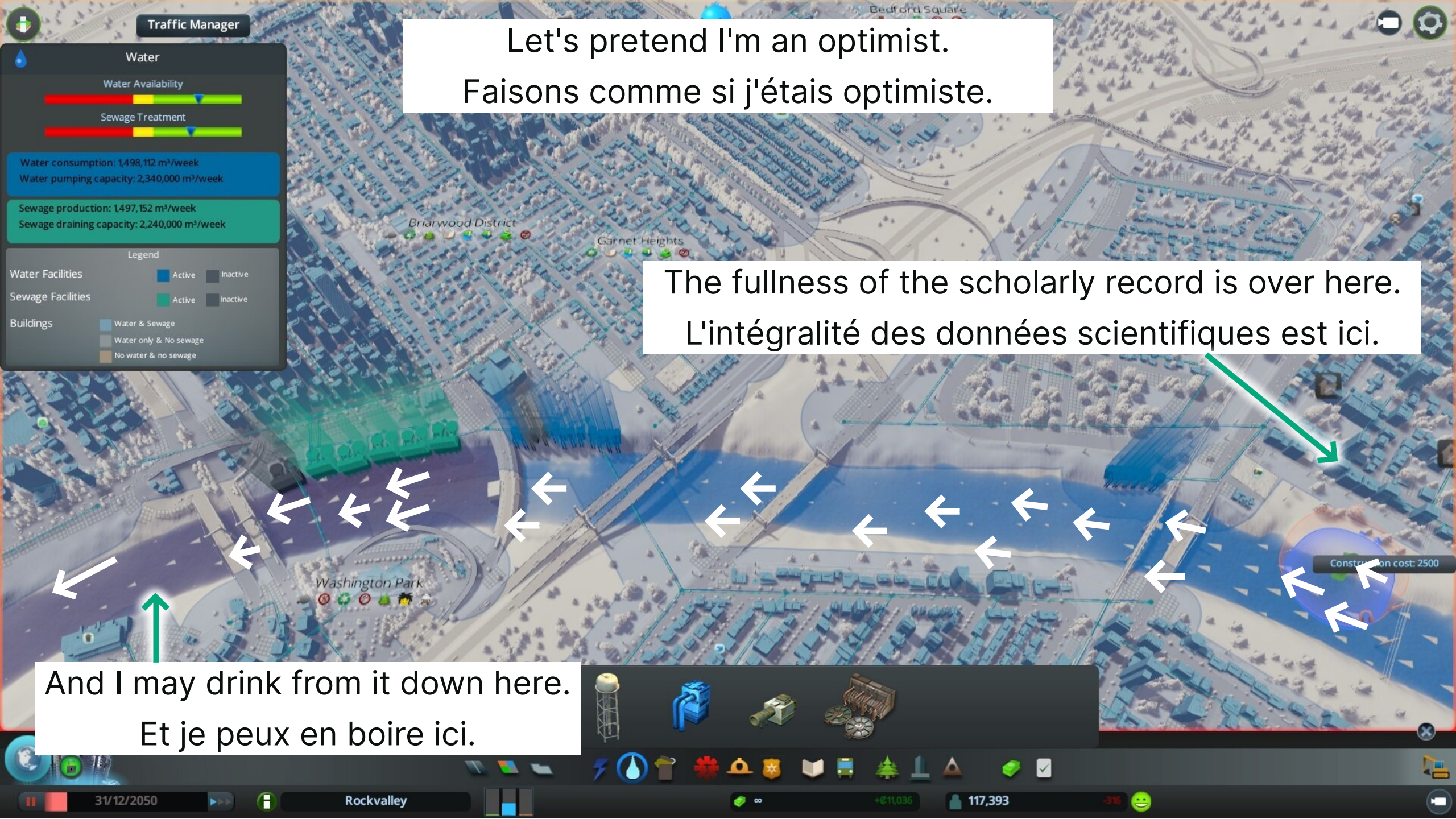
31/12/2050

117,393

+€11,036

3%



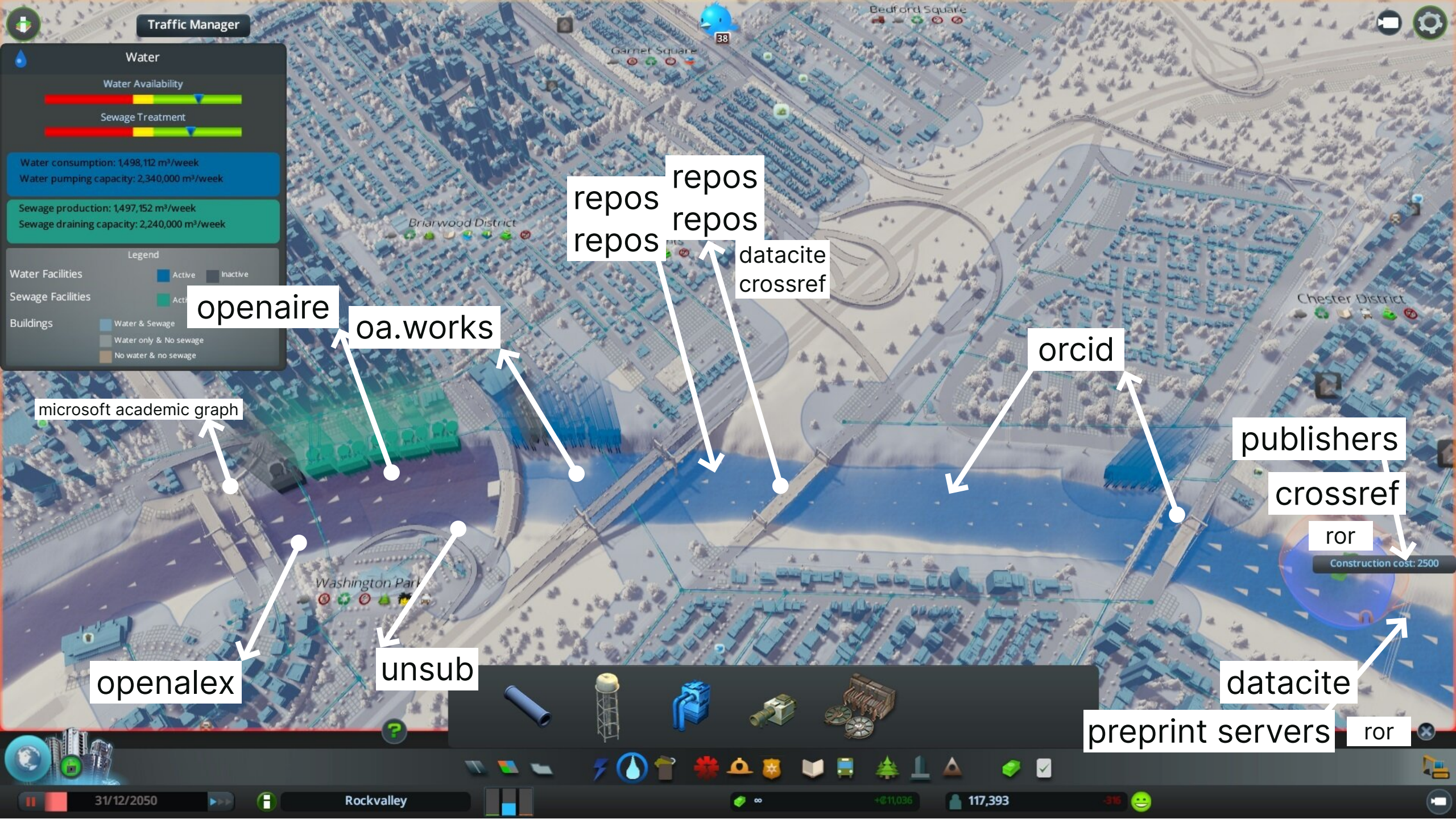


Let's pretend I'm an optimist.  
Faisons comme si j'étais optimiste.

The fullness of the scholarly record is over here.  
L'intégralité des données scientifiques est ici.

And I may drink from it down here.  
Et je peux en boire ici.

Construction cost: 2500



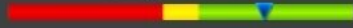
Traffic Manager

Water

Water Availability



Sewage Treatment



Water consumption: 1,498,112 m³/week  
Water pumping capacity: 2,340,000 m³/week

Sewage production: 1,497,152 m³/week  
Sewage draining capacity: 2,240,000 m³/week

Legend

- Water Facilities: Active (blue square), Inactive (grey square)
- Sewage Facilities: Active (green square)
- Buildings: Water & Sewage (light blue), Water only & No sewage (grey), No water & no sewage (brown)

openaire

oa.works

microsoft academic graph

openalex

unsub

repos  
repos  
repos

repos  
repos

datacite  
crossref

orcid

publishers

crossref

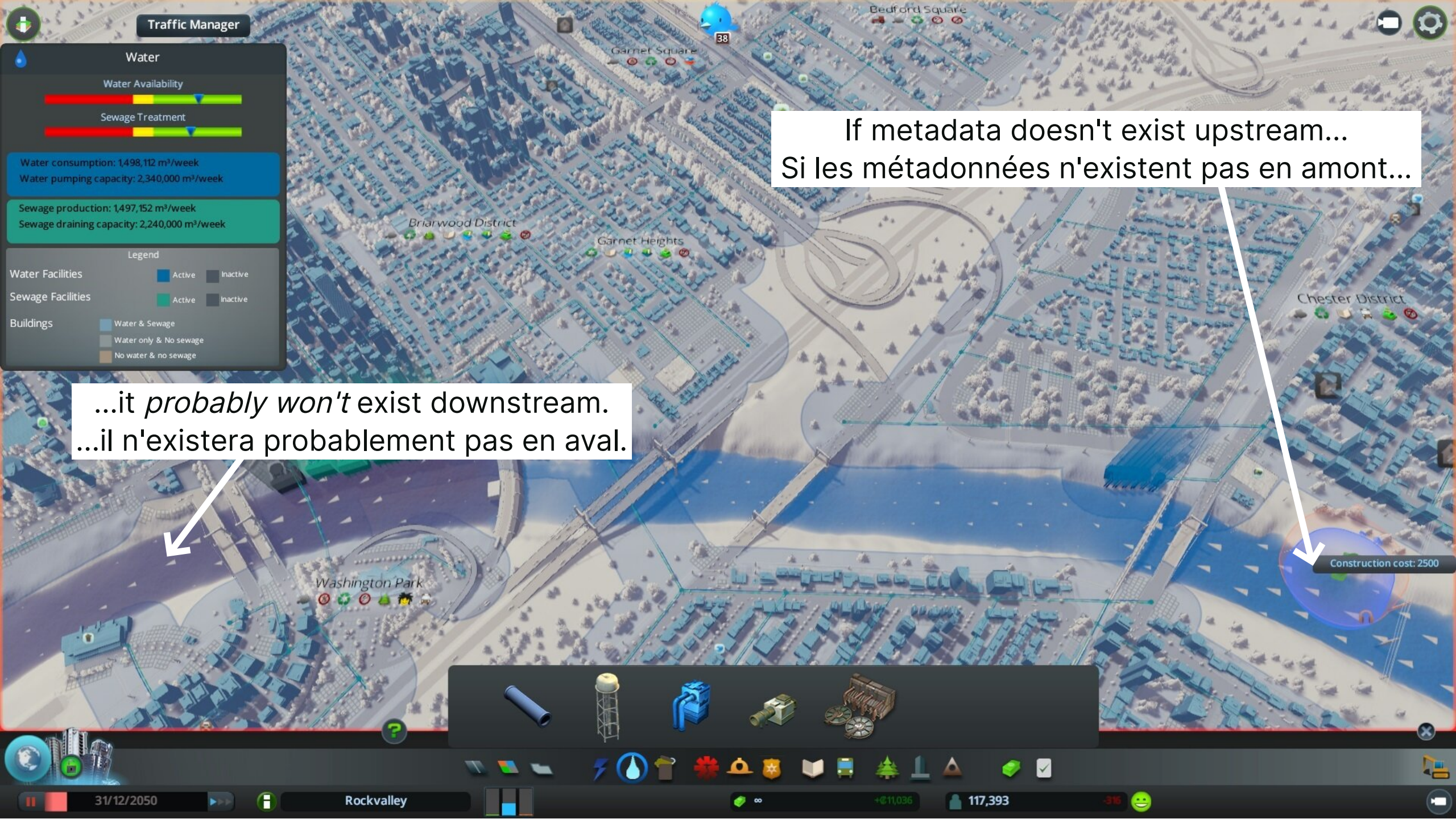
ror

Construction cost: 2500

datacite

preprint servers

ror



Traffic Manager

Water

Water Availability



Sewage Treatment



Water consumption: 1,498,112 m³/week  
Water pumping capacity: 2,340,000 m³/week

Sewage production: 1,497,152 m³/week  
Sewage draining capacity: 2,240,000 m³/week

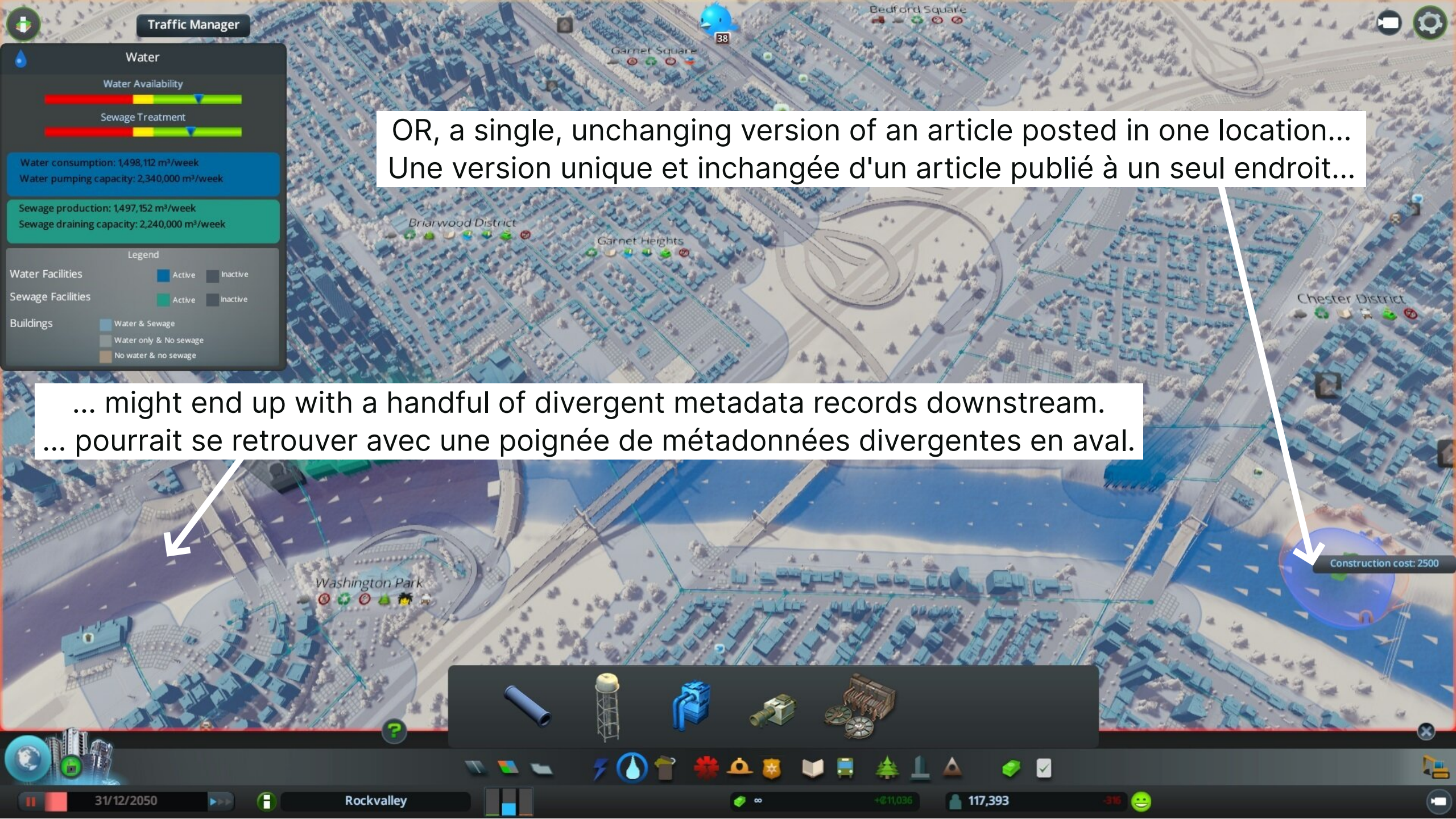
Legend

- Water Facilities: Active (blue square), Inactive (grey square)
- Sewage Facilities: Active (green square), Inactive (grey square)
- Buildings:
  - Water & Sewage (light blue square)
  - Water only & No sewage (grey square)
  - No water & no sewage (brown square)

If metadata doesn't exist upstream...  
Si les métadonnées n'existent pas en amont...

...it *probably won't* exist downstream.  
...il n'existera probablement pas en aval.

Construction cost: 2500



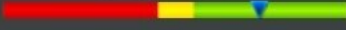
### Traffic Manager

#### Water

Water Availability



Sewage Treatment



Water consumption: 1,498,112 m³/week  
Water pumping capacity: 2,340,000 m³/week

Sewage production: 1,497,152 m³/week  
Sewage draining capacity: 2,240,000 m³/week

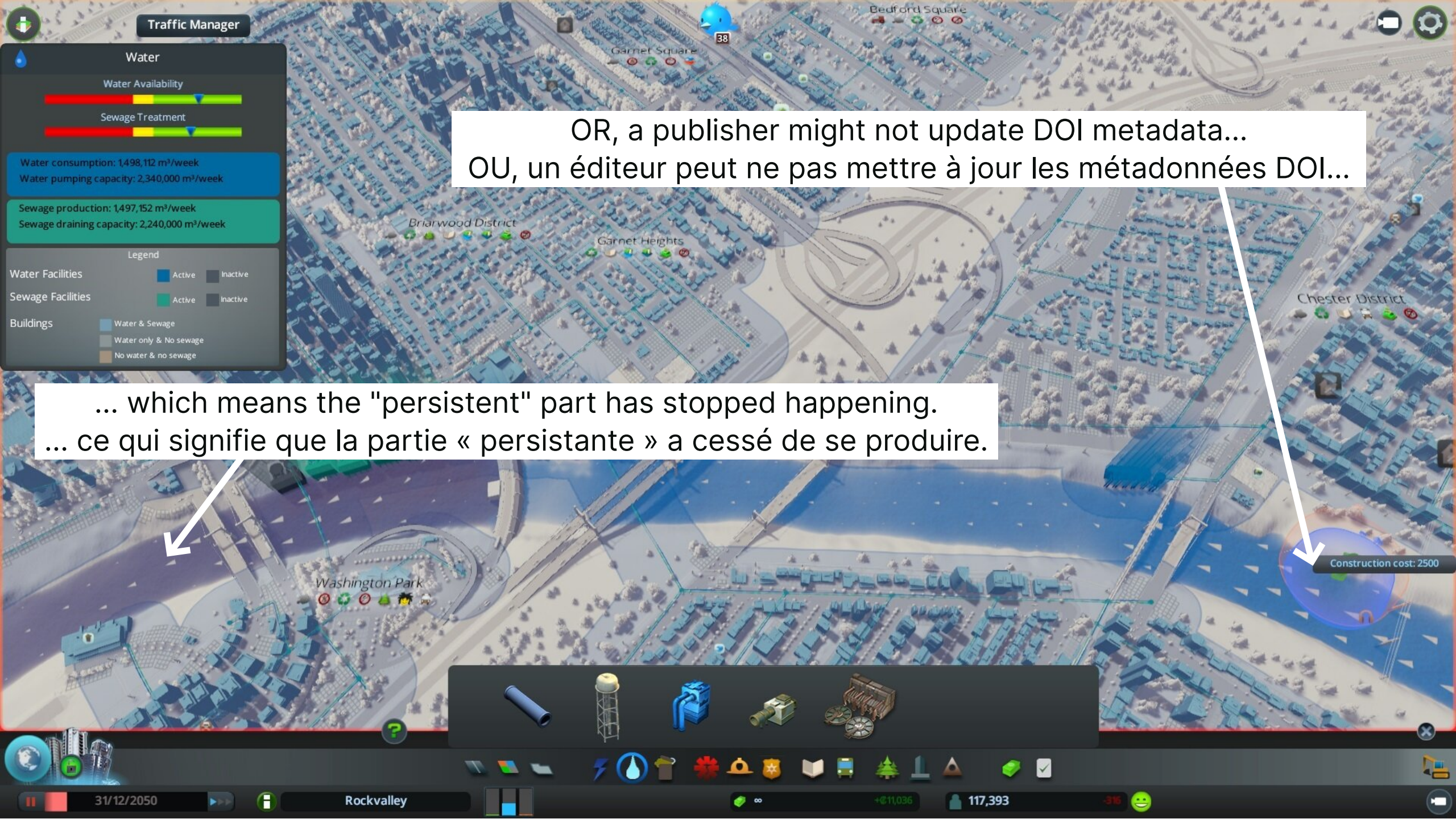
#### Legend

- Water Facilities: ■ Active ■ Inactive
- Sewage Facilities: ■ Active ■ Inactive
- Buildings:
  - Water & Sewage
  - Water only & No sewage
  - No water & no sewage

OR, a single, unchanging version of an article posted in one location...  
Une version unique et inchangée d'un article publié à un seul endroit...

... might end up with a handful of divergent metadata records downstream.  
... pourrait se retrouver avec une poignée de métadonnées divergentes en aval.

Construction cost: 2500



### Traffic Manager

#### Water

Water Availability



Sewage Treatment



Water consumption: 1,498,112 m<sup>3</sup>/week  
Water pumping capacity: 2,340,000 m<sup>3</sup>/week

Sewage production: 1,497,152 m<sup>3</sup>/week  
Sewage draining capacity: 2,240,000 m<sup>3</sup>/week

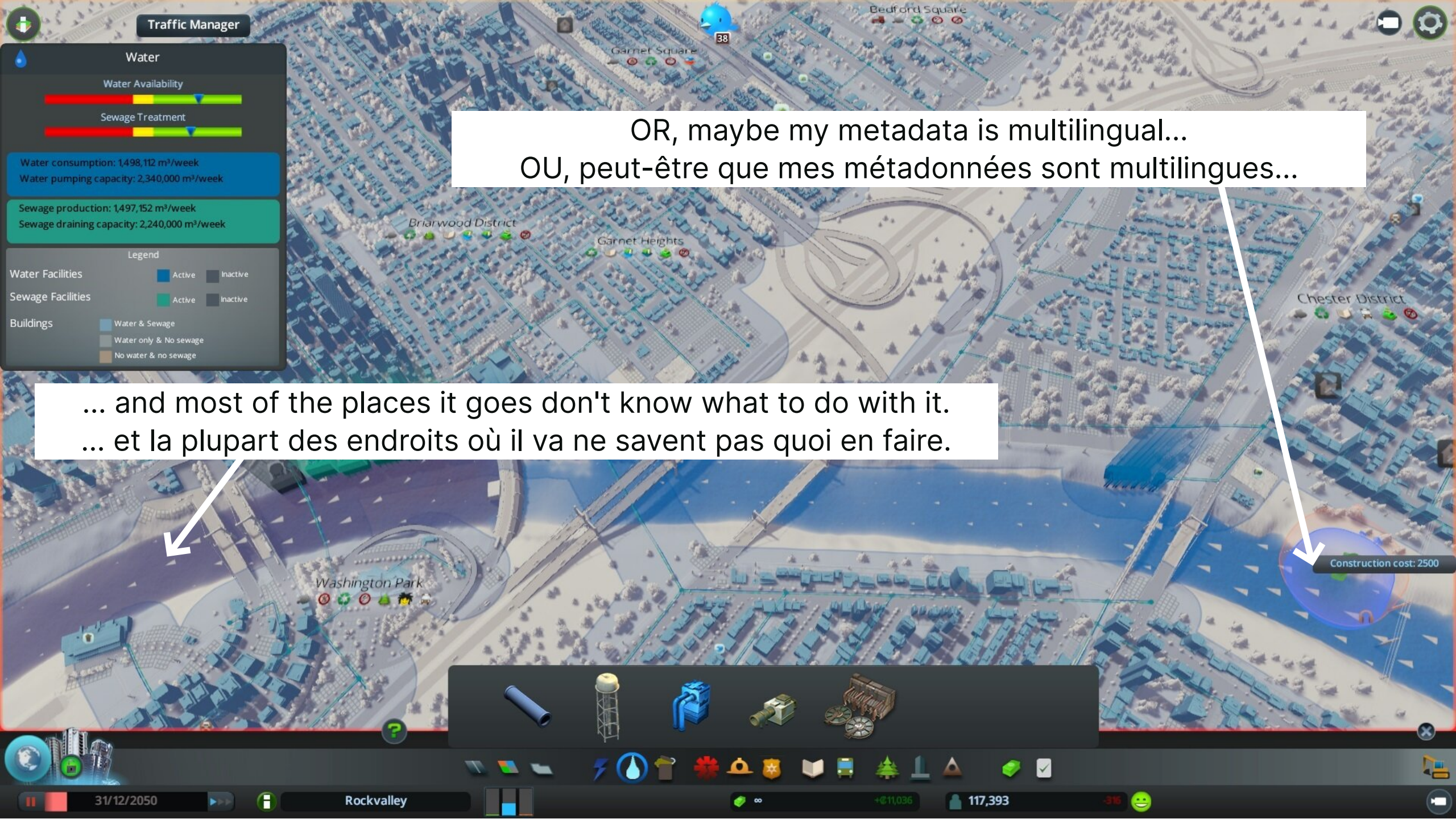
#### Legend

- Water Facilities: Active (blue square), Inactive (grey square)
- Sewage Facilities: Active (green square), Inactive (grey square)
- Buildings: Water & Sewage (light blue square), Water only & No sewage (grey square), No water & no sewage (brown square)

OR, a publisher might not update DOI metadata...  
OU, un éditeur peut ne pas mettre à jour les métadonnées DOI...

... which means the "persistent" part has stopped happening.  
... ce qui signifie que la partie « persistante » a cessé de se produire.

Construction cost: 2500



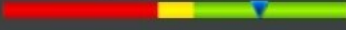
Traffic Manager

Water

Water Availability



Sewage Treatment



Water consumption: 1,498,112 m<sup>3</sup>/week  
Water pumping capacity: 2,340,000 m<sup>3</sup>/week

Sewage production: 1,497,152 m<sup>3</sup>/week  
Sewage draining capacity: 2,240,000 m<sup>3</sup>/week

Legend

- Water Facilities: Active (blue square), Inactive (grey square)
- Sewage Facilities: Active (green square), Inactive (grey square)
- Buildings: Water & Sewage (light blue), Water only & No sewage (grey), No water & no sewage (brown)

OR, maybe my metadata is multilingual...  
OU, peut-être que mes métadonnées sont multilingues...

... and most of the places it goes don't know what to do with it.  
... et la plupart des endroits où il va ne savent pas quoi en faire.

Construction cost: 2500

Washington Park

Rockvalley

31/12/2050

117,393

3%

# optimism

Optimism isn't unfounded, but there are a few reality checks we need to address here.

# optimisme

L'optimisme n'est pas infondé, mais il faut tenir compte de certaines réalités.

# compatibility

For example, **not all metadata is compatible**. The existence of a specific metadata element in, say, the Datacite schema, does not ensure the existence of a parallel metadata element in Dublin Core. Crossref's schema  $\neq$  ORCID's schema. This metadata has to be crosswalked between schemas.

# la compatibilité

Par exemple, **toutes les métadonnées ne sont pas compatibles**. L'existence d'un élément de métadonnées spécifique dans le schéma Datacite, par exemple, ne garantit pas l'existence d'un élément de métadonnées parallèle dans Dublin Core. Le schéma de Crossref  $\neq$  le schéma de l'ORCID. Ces métadonnées doivent faire l'objet d'une concordance entre les schémas.



# persistence

The other mistake we make here is in assuming metadata records are updated as our schema, tools, and interests develop over time.

Let's look at a quick example using Crossref Participation Reports.

# persistance

L'autre erreur que nous commettons ici est de supposer que les enregistrements de métadonnées sont mis à jour au fur et à mesure que notre schéma, nos outils et nos intérêts évoluent.

Prenons un exemple rapide en utilisant les rapports de participation de Crossref.

Find a member

Learn more

Richer metadata makes content useful.  
**Make sure your work can be found.**

Public Library of Science (PLOS)

**48,749**  
Total registered content items

Content type: Journal articles



Journal articles 48,749

Journal articles



Search by title



Current content



References



ORCID IDs



Funder Registry IDs



Funding award numbers



Crossmark enabled



Text mining URLs



License URLs



Similarity Check URLs



Abstracts



# PLOS Current

(the last 2 years)

Find a member

Learn more

Richer metadata makes content useful. **Make sure your work can be found.**

Public Library of Science (PLOS)

**359,711**

Total registered content items

Content type: Journal articles

Journal articles **359,704** Datasets 7

Journal articles

Search by title

All time

# PLOS All Time

(all records in Crossref)

References

93%

ORCID IDs

41%

Funder Registry IDs

31%

Funding award numbers

27%

Crossmark enabled

99%

Text mining URLs

0%

License URLs

99%

Similarity Check URLs

92%

Abstracts

19%

Richer metadata makes content useful.  
Make sure your work can be found.

Springer Publishing Company

4,694  
Total registered  
content items

Content type: Journal articles

Journal articles 1,238 Books 118 Book chapters 3,338

Journal articles

Search by title

Current content

References

77%

ORCID IDs

0%

Funder Registry IDs

0%

Funding award numbers

0%

Crossmark enabled

0%

Text mining URLs

0%

License URLs

0%

Similarity Check URLs

100%

Abstracts

89%

# Springer Current

Richer metadata makes content useful.  
Make sure your work can be found.

Springer Publishing Company

43,680  
Total registered content items

Content type: Journal articles

Journal articles 17,202 Books 950 Book chapters 25,528

Journal articles

Search by title

All time

References

40%

ORCID IDs

6%

Funder Registry IDs

0%

Funding award numbers

0%

Crossmark enabled

0%

Text mining URLs

0%

License URLs

0%

Similarity Check URLs

96%

Abstracts

63%

# Springer All Time

# ease

Outside the agencies who provide registration services, a (probably) disproportionate amount of the responsibility and labour around PIDs rests on the shoulders of researchers. We need them to get the metadata right and care about these connections so that we (and *they* can enjoy them).

But metadata is often something that *happens to* researchers more than it *happens for* researchers.

# faciliter

En dehors des agences qui fournissent des services d'enregistrement, une part (probablement) disproportionnée de la responsabilité et du travail autour des PID repose sur les épaules des chercheurs. Nous avons besoin d'eux pour obtenir des métadonnées correctes et pour se préoccuper de ces connexions afin que nous puissions en profiter (et qu'ils puissent en profiter).

Mais les métadonnées sont souvent quelque chose qui arrive aux chercheurs plus qu'elles n'arrivent pour les chercheurs.

I have spent pretty much my whole career as a librarian desperately trying to get faculty to care about things that / *believe they should care about.*

**So have you.**

What's the hit rate?

These are busy, tired people at the mercy of a predatory publishing industry.

J'ai passé pratiquement toute ma carrière de bibliothécaire à essayer de faire en sorte que les enseignants s'intéressent à des choses qui, selon moi, devraient les intéresser.

**Vous aussi.**

Quel est le taux de réussite ?

Il s'agit de personnes occupées, fatiguées et à la merci d'une industrie de l'édition prédatrice.

I promise you, I do have  
some optimism.

---

Je vous promets que je  
suis optimiste.



Persistent identifiers – and the connections they expose – aid dramatically the **discovery, dissemination, and attribution** of all these disparate products of research.

---

Les identificateurs permanents - et les connexions qu'ils révèlent - facilitent considérablement la **découverte**, la **diffusion** et **l'attribution** de tous ces produits de recherche disparates.

**We just need to understand how PIDs  
work and have sensible expectations!**

---

**Il suffit de comprendre le  
fonctionnement des PID et d'avoir  
des attentes raisonnables !**

Metadata is always going to be a little bad, and we should make peace with that.

---

Les métadonnées seront toujours un peu mauvaises, et nous devrions nous en accommoder.

within reason

---

dans les limites du raisonnable

yes, mike nason just said that.

---

oui, mike nason vient de le dire.

"Please do this so that we may better measure you", is a gross sales pitch.

---

"Faites ceci pour que nous puissions mieux vous évaluer" est un argument de vente grossier.

If I felt bullied into using ORCID, I would simply not use it. I'm like that. And it would totally be my call.

---

Si je me sentais contraint d'utiliser ORCID, je ne l'utiliserais tout simplement pas. Je suis comme ça. Et c'est moi qui déciderais.

We can pitch PIDs on their inherent value of discovery and time savings. Let's leave the cop stuff off the list.

---

Nous pouvons présenter les NID sur leur valeur inhérente de découverte et de gain de temps. Laissons de côté l'aspect policier de la liste.



A commitment to PIDs is a commitment to *persistence of service*. Many service providers aren't great at this!

---

S'engager à respecter les NID, c'est s'engager à assurer la persistance du service. De nombreux prestataires de services ne sont pas très doués pour cela !

If you're about to start using PIDs with, say, your repository, ask yourself if you will be able to maintain those endpoints.

---

Si vous êtes sur le point de commencer à utiliser des PIDs avec, par exemple, votre référentiel, demandez-vous si vous serez en mesure de maintenir ces points de terminaison.

Most metadata doesn't just happen,  
people have to create it. Metadata  
requires labour.

---

La plupart des métadonnées ne sont pas  
le fruit du hasard, elles doivent être  
créées par des personnes. Les  
métadonnées requièrent du travail.

The past is full of metadata that is as unlikely to be updated as it would be arduous and costly to update.

---

Le passé est rempli de métadonnées qui ont aussi peu de chances d'être mises à jour qu'il serait difficile et coûteux de le faire.

And many publishers are as disinterested in solving that problem as they are interested in taking your money.

---

Et de nombreux éditeurs sont aussi peu intéressés par la résolution de ce problème qu'ils le sont par le fait de vous soutirer de l'argent.

Omniscience is likely to elude us. But that doesn't mean we can't save people real time *now* with support and expertise.

---

L'omniscience nous échappera probablement. Mais cela ne signifie pas que nous ne pouvons pas sauver les gens en temps réel grâce à notre soutien et à notre expertise.

Plus, who are we kidding?  
Omniscience was never going to  
work out in *our* favour anyway.

---

Et puis, de qui se moque-t-on ?  
L'omniscience n'allait de toute  
façon jamais jouer en notre faveur.

So look. I set out today hoping to demystify the automagical. Now that we're at the end of this thing, I'm not totally confident I've done that. But, I *very much hope* I've shed some light on a subject we often gloss over when it comes to limitations.

I remain a passionate supporter of OSI. I hope this was in *any way* useful to you.

Alors, regardez. Je suis parti aujourd'hui dans l'espoir de démystifier l'automagie. Maintenant que nous sommes à la fin de cette journée, je ne suis pas totalement convaincu d'y être parvenu. Mais j'espère vraiment avoir jeté un peu de lumière sur un sujet que nous négligeons souvent lorsqu'il s'agit de limitations.

Je reste un fervent partisan des OSI. J'espère que cela vous a été utile d'une manière ou d'une autre.



Jeez, I hope this was *something*.

**Thanks** for sticking it out.

**Questions?** Demands for refunds?

J'espère que c'était quelque chose.

**Merci** d'avoir tenu bon.

**Des questions ?** Des demandes de remboursement ?