

Retour d'expérience : la "FAIRisation" de Norine

Areski Flissi¹, Maude Pupin¹, Valérie Leclère², Mickael Malandran³

¹Laboratoire CRISAL (UMR CNRS 9189, Univ. Lille, Centrale Lille)

²UMR transfrontalière BioEcoAgro (INRAE, Univ. Lille, ULiège, UPJV), Institut Charles Viollette

³ Univ Lille, Service Commun de documentation

la Fabrique de la Science Ouverte

22 novembre 2022 | 14h-16h | LILLIAD

Un projet / une collaboration pluri-disciplinaire

- ▶ Laboratoire CRISAL – UMR CNRS 9189, Univ. Lille
 - ▶ Areski Flissi – Ingénieur de Recherche en informatique,
 - ▶ Maude Pupin – Maître de conférences en informatique
 - ▶ Laurent Noé – Maître de conférences en informatique
- ▶ UMR transfrontalière BioEcoAgro (INRAE, Univ. Lille, ULiège, UPJV – Institut Charles Viollette)
 - ▶ Valérie Leclère – Professeure en microbiologie
- ▶ Plateforme Bilille, UAR 2014 US41 PLBS (Univ. Lille, CNRS, Inserm, CHU Lille, Institut Pasteur Lille)
 - ▶ Pierre Péricard – Ingénieur de Recherche en bioinformatique
- ▶ Univ Lille, Service commun de documentation
 - ▶ Marie Cros – Chargée de projets données
 - ▶ Alicia León y Barella – Cheffe du service Science ouverte
 - ▶ Mickael Malandran – Chargé de coordination Archives Ouvertes
 - ▶ Camille Vacher – Référent Science Ouverte



- ▶ Contexte : les peptides non-ribosomiques (NRP)
 - ▶ définition, intérêts des NRPs
- ▶ La plate-forme logicielle Norine
 - ▶ pourquoi faire ? pour qui ?
- ▶ **FAIRisation** de Norine
 - ▶ historique, démarche
- ▶ Bilan retour d'expérience
- ▶ Conclusion



Institut
Charles
Viollette



CRISTAL
Centre de Recherche Interdisciplinaire
Digital et Automatique de Lille



Université
de Lille



Les peptides non-ribosomiques (NRP)



Institut
Charles
Viollette



CRISTAL
Centre de Recherche Interdisciplinaire
Sémiotique et Automatique de Lille

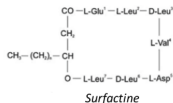
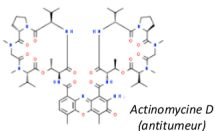
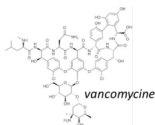


Université
de Lille

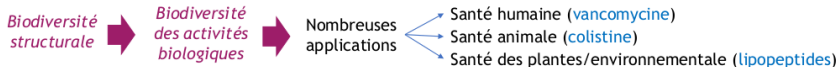


Les peptides non-ribosomiques (NRP)

- ▶ Sont des **polymères d'acides aminés (AA) et autres monomères** (> 500)
- ▶ Assemblés dans la cellule par des complexes multi-enzymatiques : NRPS (NonRibosomal Peptide Synthetases)
- ▶ Métabolites secondaires **produits par des microorganismes** (bactéries et champignons)



Les peptides non-ribosomiques (NRP)



Des données issues de nombreuses disciplines/nombreux experts :



→ Nécessité d'homogénéiser ces données

Acteurs des équilibres nécessaires à la santé globale de la planète (One Health)



Les peptides non-ribosomiques (NRP)

- ▶ L'étude des peptides non-ribosomiques nécessite des outils spécifiques
 - ▶ Pas d'outil bioinformatique existant
 - ▶ ce ne sont pas des séquences protéiques
 - ▶ ils sont plus grands que la plupart des métabolites (petites molécules du vivant)
 - ▶ pas de base de données dédiée
 - ▶ Problèmes algorithmiques intéressants
 - ▶ structures = graphes étiquetés non orientés
 - ▶ > 500 étiquettes ; max 26 noeuds avec 1 à 5 voisins
 - ▶ graphes planaires

Nous avons donc créé **Norine** l'unique plate-forme logicielle dédiée aux peptides non-ribosomiques

<http://norine.univ-lille.fr/norine>



Institut
Charles
Viollette



CRISTAL
Centre de Recherche en Informatique
Sémiotique et Automatique de Lille



Université
de Lille



La plateforme logicielle Norine

Présentation

- ▶ Projet né en 2006 [Cab+08], collaboration entre bio-informaticiens de CRISTAL et microbiologistes de l'ICV (Institut Charles Violette), co-encadrement d'une thèse
- ▶ Unique ressource pour les **peptides non-ribosomiques** (NRPs)
- ▶ **Données de Norine référencées** par plusieurs ressources internationales : PDB (*Protein Data Bank*), BIRD, MIBiG, ChemBL...

Norine	annotation search	structure search	monomers	my	REST API	Terms of use	?
--------	-------------------	------------------	----------	----	----------	--------------	---

[home](#) > [NRP](#) > [norine](#)

NORINE is a platform that includes a **database** of nonribosomal peptides together with tools for their analysis. Norine currently contains **1740 peptides**.

The name **Norine** stands for **N**onribosomal **i**peptides, with **ine** as a typical ending of peptide names.

For each **peptide**, the **database** stores its structure as well as various annotations such as the biological activity, producing organisms, bibliographical references among others. The **database** can be queried in order to search for peptides through their annotations as well as through their monomeric structures. In the latter case, the user can specify either the composition, the whole structure or a structural pattern (possibly including "undefined monomers") of the searched peptide.

To cite Norine

Norine is freely available to everybody, under the **Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International** ([CC BY-NC-SA 4.0](#))



See [terms of use](#) for more details. If you use the database, please cite:

Norine: update of the nonribosomal peptide resource. Arrosti Fissi, Emma Ricard, Clémentine Compant, Mickaël Chevalier, Yoann Duthéne, Jürg Mikiele, Philippe Jacques, Christophe Fabiani, Frédéric Linares, Valérie Leclère and Maude Pugin
Nucleic Acids Research, Nov. 2019, [glz1000](#), <https://doi.org/10.1093/nar/gkz1000>

[SiteMap](#) [News](#) [Last entries](#) [Contributors](#) [What are nonribosomal peptides](#) [Contact information](#)

News

July 2022: Award for Norine!

- Norine received the Open Science Research Data Award 2022. [More info](#)

January 2022: New release for Norine

- Norine webapp and database migration to new server. It is now hosted on the cloud of Lille University. Norine is now accessible with the new URL.

<http://norine.univ-lille.fr>



La plate-forme logicielle Norine



Institut
Charles
Viollette



CRISTAL
Centre de Recherche Interdisciplinaire
Sémiotique et Automatique de Lille



Université
de Lille

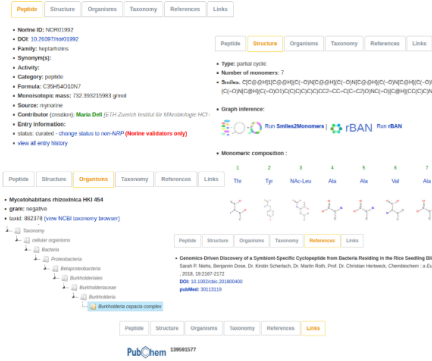


La plate-forme logicielle Norine

Présentation

- Base de données dédiée aux NRP_s et outils pour leur **analyse**, leur **recherche**, leur **comparaison** et leur **visualisation**

heptarhizin



• **Nonline ID:** NCR0193Z
• **DOI:** 10.26207/nor1990
• **Family:** heptarhizins
• **Synonym(s):**
• **Activity:**
• **Category:** peptide
• **Formula:** C₂₉H₄₂O₁₀N₂
• **Monoisotopic mass:** 752.31521983 g/mol
• **Source:** mytiline
• **Credit/Buyer (optional):** Maria Del (ETH Zürich Institut für Mikrobiologie HCI)
• **Entry information:**
• **status:** curated - (change status to non-NRP (Narine validators only))
• **view all entry history**

• **Mycohitobolans rhizotinctica HMI-854**
• **gnwl:** negative
• **taxid:** 822179 (view NCBI taxonomy browser)

Taxonomy

- Kingdom: cellular organisms
- Phylum: Basidiomycota
- Class: Basidiomycetes
- Order: Agaricomycetales
- Family: Boletaceae
- Genus: Mycohitobolans
- Species: Mycohitobolans rhizotinctica

Monomer composition:
1 Thr 2 Tyr 3 NMe-Leu 4 Ala 5 Ala 6 Val 7 Ala

Graph representation: Thr,Tyr,NMe-Leu,Ala,Ala,Val,Ala @9.2,1 @0.4 @0 @6.5 @1.6 @0.3 @3.4

Atomic structure:

• **Genomics Driven Discovery of a Substituted Specific Cyclicpeptide from Dacteria Residing in the Free Seaweed Stiglit Fungus**
Sarah P. Math, Bergesen Deok, Dr. Kavin Srichewat, Dr. Marit Math, Prof. Dr. Christian Höferbach, Christof Buchen: - European Journal of chemical Biology
2023, 19:1249-1272
DOI: 10.1002/ejcb.202300440
pubMed: 39113119

Pubchem 139593577

Logo: BioEcoAgro, Institut Charles Viollette, CRISTAL, CNRS, Université de Lille

La plateforme logicielle Norine

Présentation

- **Recherche**, via des interfaces dynamiques, par **annotations** (nom, famille, activités biologiques, catégorie, organismes producteurs, références, etc.)

Norine



home > NRP > norine

New! You can now create sophisticated queries, by using this flexible and dynamic form.

Search form with the following criteria:

- Status: curated
- AND Smiles: Search only peptides with smiles defined
- AND Number of monomers: > 10
- AND Source: norine
- AND NOT Category: glycopeptide lipopeptide peptaibol chromosomeptide peptide PK-NRP

submit



Search for:

(status = "curated") AND (Peptides that smiles is defined) AND (source is "norine") AND (containing a number of monomers SUP 10) AND NOT (category is glycopeptide | chromosomeptide)

[Back to the form](#)

88 peptide(s), grouped in 24 famille(s)

Select/unselect all

Calcium dependent antibiotic (2)

- CDA-Ib
- CDA-Ia

actinoidin (1)

- actinoidin.A

amphysin (5)

- isosin
- amphysin
- phyllopeptin
- bersin
- arthrofactin

- ▶ **Recherche par structure** (comparaison de graphes [Cab+09])
 - ▶ recherche de sous-motifs dans un graphe (*pattern*), exemple :
"Ala, X, [Ser | Gly]@1@0, 2@1"

Structure-based search [?]

peptide(s) containing the structural pattern: Ala.Gly.X@2@2@0,1

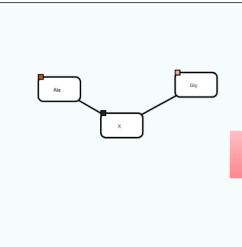
Open the graphical editor [?]

Search monomers

- X
- ▲ Asp*
- ▼ Asp*
- ▲ AB*
- ▼ AB*
- ▲ Ala+Gly
- ▼ Ala*
- ▼ Gly*
- ▲ Abx
- ▼ Asp*
- ▼ Asp*
- ▲ Bst*
- ▼ Bst*
- ▲ Bt*
- ▼ Bt*
- ▲ Cap*
- ▼ Cap*

Active Monomer(s)

Redraw Clear Show .png image



Search for:

Peptide(s) containing at least one of these fragments with no less than 0 monomer :

1. Ala.Gly.X@2@2@0,1

27 peptides found.

Select/Unselect all

peptide	download
antismad11r	<input checked="" type="checkbox"/>
antismad11s	<input checked="" type="checkbox"/>
apikmglyc1	<input checked="" type="checkbox"/>
apikmglyc2	<input checked="" type="checkbox"/>
apikmglyc3	<input checked="" type="checkbox"/>
apikmglyc4	<input checked="" type="checkbox"/>
apikmglyc5	<input checked="" type="checkbox"/>
apikmglyc6	<input checked="" type="checkbox"/>
apikmglyc7	<input checked="" type="checkbox"/>
apikmglyc8	<input checked="" type="checkbox"/>
apikmglyc9	<input checked="" type="checkbox"/>
apikmglyc10	<input checked="" type="checkbox"/>
apikmglyc11	<input checked="" type="checkbox"/>
apikmglyc12	<input checked="" type="checkbox"/>
apikmglyc13	<input checked="" type="checkbox"/>
apikmglyc14	<input checked="" type="checkbox"/>
apikmglyc15	<input checked="" type="checkbox"/>
apikmglyc16	<input checked="" type="checkbox"/>
apikmglyc17	<input checked="" type="checkbox"/>
apikmglyc18	<input checked="" type="checkbox"/>
apikmglyc19	<input checked="" type="checkbox"/>
apikmglyc20	<input checked="" type="checkbox"/>
apikmglyc21	<input checked="" type="checkbox"/>
apikmglyc22	<input checked="" type="checkbox"/>
apikmglyc23	<input checked="" type="checkbox"/>
apikmglyc24	<input checked="" type="checkbox"/>
apikmglyc25	<input checked="" type="checkbox"/>
apikmglyc26	<input checked="" type="checkbox"/>
apikmglyc27	<input checked="" type="checkbox"/>
apikmglyc28	<input checked="" type="checkbox"/>
apikmglyc29	<input checked="" type="checkbox"/>
apikmglyc30	<input checked="" type="checkbox"/>
apikmglyc31	<input checked="" type="checkbox"/>
apikmglyc32	<input checked="" type="checkbox"/>
apikmglyc33	<input checked="" type="checkbox"/>
apikmglyc34	<input checked="" type="checkbox"/>
apikmglyc35	<input checked="" type="checkbox"/>
apikmglyc36	<input checked="" type="checkbox"/>
apikmglyc37	<input checked="" type="checkbox"/>
apikmglyc38	<input checked="" type="checkbox"/>
apikmglyc39	<input checked="" type="checkbox"/>
apikmglyc40	<input checked="" type="checkbox"/>
apikmglyc41	<input checked="" type="checkbox"/>
apikmglyc42	<input checked="" type="checkbox"/>
apikmglyc43	<input checked="" type="checkbox"/>
apikmglyc44	<input checked="" type="checkbox"/>
apikmglyc45	<input checked="" type="checkbox"/>
apikmglyc46	<input checked="" type="checkbox"/>
apikmglyc47	<input checked="" type="checkbox"/>
apikmglyc48	<input checked="" type="checkbox"/>
apikmglyc49	<input checked="" type="checkbox"/>
apikmglyc50	<input checked="" type="checkbox"/>

► Analyse et filtrage des données de Norine (histogrammes, camemberts...)



Search for:

(status = 'curated') AND (Peptides that smiles is defined) AND (source is 'norine').
is peptaibol | chromopeptide)



68 peptide(s), grouped in 24 famille(s)

Select/Unselect all

Calcium dependent antibiotic (2)

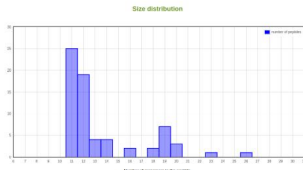
- CDA4b
- CDA4a

actinoidin (1)

- actinoidin A

amphisin (5)

- lakisin
- amphisin
- pholipepín
- tensin
- arthrofactin



La FAIRisation de Norine



Institut
Charles
Viollette



CRISTAL
Centre de Recherche Interdisciplinaire
Sémiotique et Automatique de Lille



Université
de Lille



La FAIRisation de Norine

Historique

- ▶ Principes de **libre accès**, d'**interopérabilité** et de **réutilisation** dès la naissance du projet en 2006!
 - ▶ <http://norine.univ-lille.fr> pour l'accès aux entrées de Norine (+ de 1000 NRPs)
 - ▶ description fine de chaque molécule
 - ▶ téléchargements possibles dans différents formats (TXT, CSV, XML, HTML)

—> *Pratiques communément appliquées depuis de nombreuses années en bio-informatique*



La FAIRisation de Norine

Historique

- ▶ Ouverture en **2015** à l'*expert sourcing* : déploiement de **MyNorine** [Fli+15]
 - ▶ + de **100** utilisateurs experts enregistrés
 - ▶ ~**100** contributions (nouvelles soumissions de NRPs ou d'annotations ou modifications/corrections) à ce jour
 - ▶ validation manuelle des soumissions ⇒ **préservation de la qualité des données Norine**



Institut
Charles
Viollette



CRISTAL
Centre de Recherche Interdisciplinaire
en Systèmes d'Automatique de Lille



Université
de Lille



La FAIRisation de Norine

Historique

- ▶ A partir de **2016** : accentuation de l'**interopérabilité** avec le développement d'une **API web** pour Norine (*RESTful web services*)
 - ▶ **accès aux données** de la base sous différents **formats standards** (XML/JSON) de manière *programmatique*
 - ▶ interactions avec les ressources externes

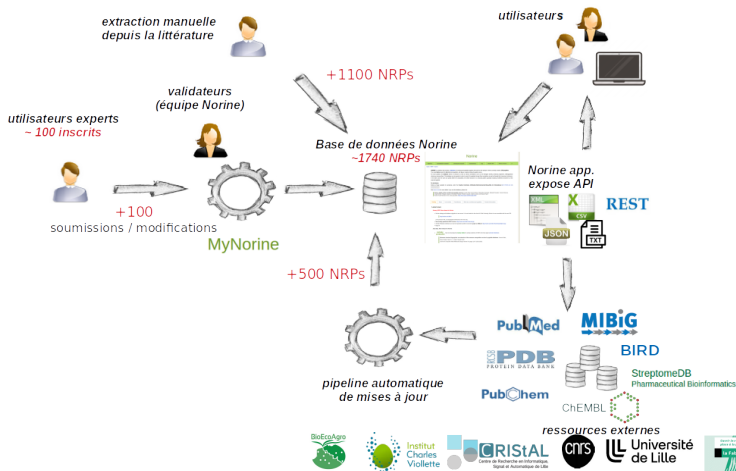
```
{
  "peptide" : {
    "description" : {
      "id" : "NOR00123",
      "name" : "[Dha7]MCRYST-E(OMe)E(OMe)",
      "activities" : ["antibiotic", "antiinflammatory"],
      "family" : "microcystin",
      "status" : "curated"
    },
    "structure" : {
      "type" : "cyclic",
      "graph" : "D-Ala, Me0-Glu, D-bMe-Asp, Me0-Glu, Adda, D-Glu, dh-Ala@1, 6@0, 2@1, 3@2, 4@3, 5@4, 6@0, 5"
    },
    "reference" :
    [
      {
        "title" : "Seven new microcystins possessing two L-glutamic acid units, isolated from Anabaena sp. strain 186",
        "authors" : [ "Sivonen K", "Namikoshi M", "Carmichael WR" ],
        "year" : "someYear",
        "journal": "Chemical research in toxicology",
        "pages" : "Feb,11(2):143-9."
      }
    ],
    "organism" : [ {
      "name" : "Anabaena",
      "taxonomy" : "cellular organisms, Bacteria, Cyanobacteria, Nostocales, Nostocaceae"
    } ]
  }
}
```



La FAIRisation de Norine

Historique

- ▶ **2019** : Déploiement d'une **mise à jour majeure (+500 NRPs)** [Fli+19]
 - ▶ *pipeline* semi-automatique de fouilles et extractions de données à partir de sources externes



- ▶ Labellisation 2019 **ELIXIR-FR** *Service Delivery Plan* (SDP)



- ▶ **bio.tools** registry: <https://bio.tools/NORINE>

The French SDP aims at identifying databases, tools, IT infrastructures and training services that are important for the French community working in bioinformatics for life science supported or not by IFB's current nodes. In addition, these tools should be aligned with the recommendations for the best practices of ELIXIR's platforms, in on **Open Science** dynamic and reasonably sustainable.



La FAIRisation de Norine

Historique

- ▶ À partir de mi-2019 : **début d'une collaboration** avec l'équipe "données" du service Science Ouverte du *Service Commun de Documentation* de l'Université de Lille
 - ▶ Objectifs : **formalisation** de notre pratique de la **science ouverte** et **poursuite** de la démarche **FAIR**
 - ▶ Comment : organisation de plusieurs réunions de travail afin de définir et mettre en œuvre la démarche



- Attribution automatique d'un DOI (*Data Object Identifier*) à chaque NRP de Norine ⇒ pérennité des URLs

The screenshot shows the DataCite website homepage. At the top, there is a navigation bar with links for 'About us', 'Services', 'Resources', 'Community', and 'Become a member'. The main banner features the text 'WELCOME TO DATA CITE' and a sub-headline: 'Locate, identify, and cite research data, publications, and grab a DOI for research.' Below the banner are four circular icons with corresponding text: 1. A magnifying glass icon: 'Find what you're looking for by searching millions of records with extensive, reliable metadata.' 2. A circular arrow icon: 'Share your data and reuse the data of others to create the highest impact in the research community.' 3. A double checkmark icon: 'Cite your research sources with confidence, and receive proper credit when your work is reused.' 4. A link icon: 'Connect your research - publications, datasets, software, authors, institutions, and funding data all in one place.' Below this is a section titled 'Get started with DataCite!' containing three boxes: 'DataCite' (Search our registry to find datasets, software, images, and other research material), 're3data.org' (Find an appropriate repository to access and deposit research data with re3data.org), and 'Crossref' (Generate your references automatically with our easy-to-use citation formatting tool).

- <https://doi.org/10.26097/nor00342> ⇒ norine.univ-lille.fr/norine/result.jsp?ID=NOR00342



► Attribution automatique d'un DOI

- (i) génération du fichier XML Datacite qui contient les méta-données du NRP
- (ii) envoi des méta-données à l'aide d'une requête **HTTP POST**,
- (iii) enregistrement de l'URL à l'aide d'un **HTTP PUT**
 - ex. avec les données suivantes :
doi=10.0341/nor00434
<https://norine.univ-lille.fr/norine/result.jsp?ID=NOR00434>

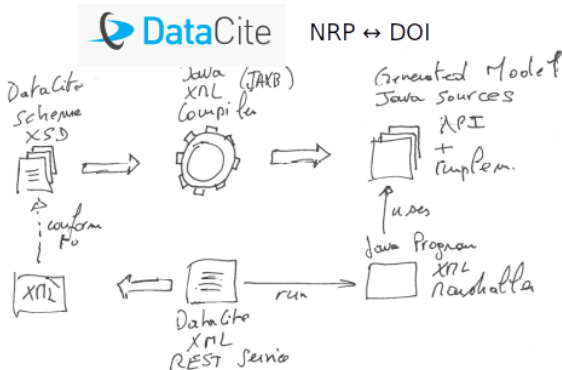


- ▶ La génération des méta-données : ex. de fichier XML conforme au schéma XML Datacite

```
<?xml version="1.0" encoding="UTF-8" ?>
<resource xmlns="http://datacite.org/schema/kernel-4" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xsi:schemaLocation="http://datacite.org/schema/kernel-4
http://schema.datacite.org/meta/kernel-4.2/metadata.xsd">
  <identifier identifierType="DOI">10.26097/nor00001</identifier>
  <creators>
    <creator>
      <creatorName nameType="Organizational">Norine</creatorName>
      <affiliation>
        Centre de Recherche en Informatique, Signal et Automatique de Lille (CRISTAL), UMR CNRS 9189
      </affiliation>
      <affiliation>Université de Lille</affiliation>
    </creator>
    <creator>
      <creatorName nameType="Organizational">CNRS (Centre national de la recherche scientifique)</creatorName>
      <affiliation>Inria</affiliation>
    </creator>
  </creators>
  <titles>
    <title>daptonycin</title>
    <title titleType="AlternativeTitle">daptonycin</title>
  </titles>
  <publisher>Norine</publisher>
  <publicationYear>2006</publicationYear>
  <resourceType resourceTypeGeneral="PhysicalObject">Peptide</resourceType>
  <contributors>
    <contributor contributorType="ContactPerson">
      <contributorName>Norine Team</contributorName>
      <affiliation>CRISTAL (UMR CNRS 9189), CRISTAL, France</affiliation>
    </contributor>
    <contributor contributorType="ContactPerson">
      <contributorName>Charles Viollette Institute, ProBioGEM team, Lille, France</contributorName>
      <affiliation>University of Lille, France</affiliation>
    </contributor>
    <contributor contributorType="ContactPerson">
      <contributorName>MICKAEL CHEVALIER</contributorName>
      <affiliation>Charles Viollette Institute, ProBioGEM team, Lille, France</affiliation>
      <affiliation>University of Lille, France</affiliation>
    </contributor>
    <contributor contributorType="ContactPerson">
      <contributorName>Emma Ricart Altiras</contributorName>
      <affiliation>Swiss Institute of Bioinformatics, SIB</affiliation>
    </contributor>
  </contributors>
  <dates>
    <date dateType="Created">2006-06-13</date>
    <date dateType="Updated">2018-12-01</date>
  </dates>
  <language>en-US</language>
  <alternateIdentifiers>
    <alternateIdentifier alternateIdentifierType="Norine">NOR00001</alternateIdentifier>
  </alternateIdentifiers>
  <formats>
    <format>application/xml</format>
  </formats>
  <version>4.2</version>
  <rightsList>
    <rights/>
  </rightsList>
  <descriptions>
    <description descriptionType="Abstract">
      <description>
        NORINE is a platform that includes a database of nonribosomal peptides together with tools for their analysis. Norine currently contains more than 1000 peptides.
      </description>
    </description>
  </descriptions>
</resource>
```



- La génération des méta-données – un (tout petit) peu de technique



► Data Licensing



► Dépôt mapreuve.com : données, schéma et documentation



► Mise en conformité **RGPD**

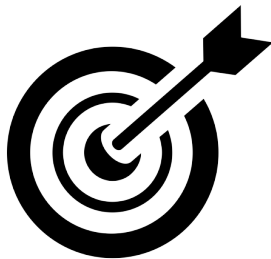


La FAIRisation de Norine

Démarche

- ▶ Plate-forme hébergée jusqu'en 2021 sur un serveur de l'équipe **Bonsai** (<https://www.cristal.univ-lille.fr/bonsai>)
- ▶ **Migration** vers un serveur du mésocentre (VM) de l'Université de Lille
 - ▶ amélioration de la disponibilité du *service*, délégation des mises à jour de sécurité
 - ▶ architecture matérielle et logicielle plus adaptée (puissance, stockage, RAM, etc.)
 - ▶ architecture 3 tiers / 4 tiers : *frontend* (pages statiques) + *backend* (applications + bdd), *dockerization*
 - ▶ <https://bioinfo.cristal.univ-lille.fr/norine> ⇒ <http://norine.univ-lille.fr>





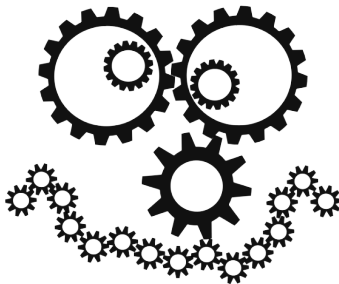
EXPÉRIMENTER



Institut
Charles
Viollette



- ▶ Prise en charge de l'abonnement à DataCite
 - ▶ D'un point de vue administratif
<https://doi.org/10.26097/nor01257>
 - ▶ D'un point de vue financier
 - ▶ D'un point de vue professionnel



- ▶ Développement d'une expertise technique
 - ▶ API
 - ▶ Schéma de méta-données
- ▶ Réalisation d'un *benchmark* sur les licences, les conditions de réutilisation et la *citabilité* des données issues de BDD de recherche

PRECONISATIONS



Bilan retour d'expérience



Institut
Charles
Viollette



CRISTAL
Centre de Recherche Interdisciplinaire
Sémiotique et Automatique de Lille



Université
de Lille

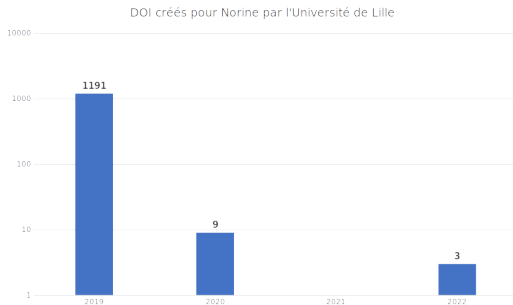


La FAIRisation de Norine

Bilan retour d'expérience

► F (*Findable*)

- base de données librement accessible <http://norine.univ-lille.fr/norine>
- DOI associé à chaque molécule
- référencée par les ressources externes telles que PDB, antiSMASH, BIRD ou MIBiG, dont les entrées pointent directement vers les données de Norine lorsque cela est pertinent.



Institut
Charles
Viollette



CRISTAL
Centre de Recherche Interdisciplinaire
en Systèmes d'Automatique de Lille



Université
de Lille



La FAIRisation de Norine

Bilan retour d'expérience

- ▶ A (*Accessible*)
 - ▶ protocoles standards du Web, libres et ouverts ([HTTP/HTTPS](#))
 - ▶ authentification nécessaire uniquement pour les contributeurs à la base de données, qui peuvent se créer un compte via l'application [MyNorine](#)



Institut
Charles
Violette



CRISTAL
Centre de Recherche Interdisciplinaire
en Systèmes d'Automatique de Lille



Université
de Lille



- ▶ I (*Interoperable*)
 - ▶ données accessibles via des interfaces web dynamiques (recherche par annotations ou structure)
 - ▶ téléchargeables sous +sieurs formats (TXT, CSV, XML, JSON...)
 - ▶ API web basée sur l'architecture REST afin de permettre une interopérabilité avec les autres bases de données



- ▶ R (*Reusable*)
 - ▶ Licence Creative Commons Attribution- NonCommercial-ShareAlike 4.0 International ([CC BY-NC-SA 4.0](#))
 - ▶ historique et traçabilité (*dates de création, contributeurs, ajouts d'annotations, modifications et commentaires, suppression...*)



Institut
Charles
Viollette



CRISTAL
Centre de Recherche Interdisciplinaire
Digital et Automatique de Lille



Université
de Lille



La FAIRisation de Norine

Bilan retour d'expérience

- ▶ Nécessite des moyens humains pour la mise en œuvre effective de **FAIR**
- ▶ Travail (d'Ingénieur au minimum) continu et sur le long terme
 - ▶ ex. attribution DOIs
- ▶ Nombreux avantages à l'"ouverture" des données de la recherche :
 - ▶ visibilité, ↑ des entrées et contributions
 - ▶ enrichissement et amélioration continue des données
 - ▶ publications : ↑ citations
 - ▶ crédibilité : ↑ aspect données ouvertes lors du dépôt de dossiers ANR!
- ▶ **Pas** d'inconvénient
 - ▶ ∅ vol de données si licence et conditions d'utilisation claires



La FAIRisation de Norine

Bilan retour d'expérience

- ▶ Points d'améliorations
 - ▶ méta-données pour le *findable*
 - ▶ disponibilité de service ? comment gérer les **nécessaires** mises à jour de l'infrastructure du mésocentre...
 - ▶ validation des contributions afin de conserver la qualité des nouvelles entrées de la base de données



Institut
Charles
Viollette



CRISTAL
Centre de Recherche Interdisciplinaire
Digital et Automatique de Lille



Université
de Lille



Côté Service Commun de la Documentation

► Points d'améliorations

- Le *schema* de Datacite est nécessairement générique, appliqué à des données qui ne le sont pas
- Vers une amélioration de la prise en compte du web des données ?

```
</creators>
</creators>
</titles>
<title daptomycin</title>
<title titleType="AlternativeTitle">daptomycin</title>
</titles>
<publisher>Norine</publisher>
<publicationYear>2006</publicationYear>
<resourceType resourceTypeGeneral="PhysicalObject">Peptide</resourceType>
<contributors>
  <contributor contributorType="ContactPerson">
    <contributorName>Norine Team</contributorName>
    <affiliations>CRISTAL (UMR CNRS 9189), ex-LIFL, France</affiliations>
  </contributor>
  <contributor contributorType="ContactPerson">
    <contributorName>Charles Viollette Institute, ProBioGEM team, Lille, France
    </affiliation>
    <affiliations>University of Lille, France</affiliations>
  </contributor>
  <contributor contributorType="ContactPerson">
    <contributorName>MICKAEL CHEVALIER</contributorName>
    <affiliations>
      Charles Viollette Institute, ProBioGEM team, Lille, France
    </affiliation>
    <affiliations>University of Lille, France</affiliations>
  </contributor>
  <contributor contributorType="ContactPerson">
    <contributorName>Emma Ricart Altiniras</contributorName>
    <affiliations>Swiss Institute of Bioinformatics, SIB</affiliations>
  </contributor>
</contributors>
<dates>
  <date dateType="Created">2006-06-13</date>
  <date dateType="Updated">2018-12-01</date>
</dates>
<language>en-US</language>
<alternateIdentifiers>
  <alternateIdentifier alternateIdentifierType="Norine">NOR00001</alternateIdentifier>
</alternateIdentifiers>
<formats>
  <format>application/xml</format>
</formats>
<version>4.2</version>
<rightsList>
  <rights/>
</rightsList>
<descriptions>
  <description descriptionType="Abstract">
    NORINE is a platform that includes a database of nonribosomal peptides together with tools for their analysis. Norine currently contains more than 1000 peptides.
  </description>

```



La FAIRisation de Norine

Bilan retour d'expérience / perspectives

Côté Service Commun de la Documentation

- ▶ Clarification du rôle du SCD : **facilitateur**
- ▶ Apport d'une expertise, mais la décision reste celle des scientifiques
- ▶ Vers un élargissement de l'expérimentation de la FAIRisation des bases de données



Institut
Charles
Violette



CRISTAL
Centre de Recherche Interdisciplinaire
Digital et Automatique de Lille



Université
de Lille



La FAIRisation de Norine

Back to the future 2008–2022

The screenshot displays the Norine database interface. It features a central graph with nodes and edges, and several surrounding panels labeled (a) through (f). Panel (a) shows a list of products. Panel (b) shows a list of suppliers. Panel (c) shows a list of products. Panel (d) shows a list of products. Panel (e) shows a list of products. Panel (f) shows a list of products.

The screenshot displays the PubChem interface for heptarhizin. It shows the chemical structure, name, and various properties. The interface includes tabs for Properties, Synonyms, Disposition, Toxicology, References, and Links. The main content area lists various identifiers and properties, including the chemical structure, name, and various identifiers.

The screenshot displays a network graph visualization. It shows a central node connected to several other nodes, forming a network. The nodes are labeled with chemical structures and names. The graph is presented in a window with a toolbar and a sidebar.



Bilan scientifique

- ▶ +2000 requêtes / mois
- ▶ 16 articles scientifiques publiés dans des journaux internationaux
 - ▶ 3 dans *Nucleic Acids Research DataBase Issue* (très sélectif)
 - ▶ co-signés avec la Belgique, le Danemark, la Suisse
 - ▶ > 600 citations cumulées
- ▶ 2 thèses en informatique et 1 en biologie co-encadrées, collaboration sur 4 autres entre CRISAL et ICV
- ▶ Collaborations internationales dont participation au CST (Belgique) et jury (Suisse) de 2 thèses
- ▶ Co-organisation de 4 workshops internationaux (2013, 2015 → Lille, 2018, 2022 → Copenhague) pour former les utilisateurs
- ▶ Participation à *HELM community* de la Pistoia Alliance, avec MERCK, Pfizer, GSK...



Remerciements

- ▶ Toute l'équipe du SCD de l'Université de Lille!
- ▶ Romain Feret (**Chargé de mission open access** et données de la recherche à l'Université de Lille)
- ▶ Myriam Caudrelier (**Chargée de la valorisation de la recherche et de l'ingénierie de projet** à CRISStAL)
- ▶ Elodie Raynal (**Juriste** à l'Université de Lille)
- ▶ Sylvain Karpf (**Chargé de valorisation** à l'Université de Lille)



Conclusion

- ▶ Une plate-forme logicielle reconnue dans le domaine de la bio-informatique **ouverte** dès la naissance du projet
- ▶ Libre **accès** aux données, **interopérabilité**, **ré-utilisabilité** *dans l'ADN* de l'équipe depuis plusieurs années
- ▶ Ingénierie logicielle CI/CD (*Continuous Integration / Continuous Delivery*)
- ▶ *Formalisation* **FAIR** grâce à une collaboration avec l'équipe "données" de l'Université de Lille
- ▶ Bilan plus que positif (mise en œuvre et résultats)



Conclusion

- ▶ Rendre la **science ouverte des données de la recherche** \Rightarrow uniquement des avantages!
- ▶ Nécessité de personnels d'appui (et les compétences techniques et du temps...) pour la mise en œuvre effective



Institut
Charles
Viollette



CRISTAL
Centre de Recherche Interdisciplinaire
Signal et Automatique de Lille



Université
de Lille



Cited publications I

- [Cab+08] S. Caboche et al. “NORINE: a database of nonribosomal peptides”. In: *Nucl. Acids Res.* 36 (2008), pp. D326–D331.
- [Cab+09] S. Caboche et al. “Structural pattern matching of nonribosomal peptides”. In: *BMC structural biology* 9.1 (2009), p. 15.
- [Fli+15] A. Flissi et al. “Norine, the knowledgebase dedicated to non-ribosomal peptides, is now open to crowdsourcing”. In: *Nucl. Acids Res.* 44 (2015), pp. D1113–D1118.
- [Fli+19] A. Flissi et al. “Norine: update of the nonribosomal peptide resource”. In: *Nucl. Acids Res. database issue* (2019).

