



www.cnrs.fr

Expérimentation Web Sémantique du projet ISTEEX

– Nicolas Thouvenin

 *thouveni*





“Construire le socle de la bibliothèque scientifique numérique nationale”



MINISTÈRE
DE L'ENSEIGNEMENT SUPÉRIEUR
ET DE LA RECHERCHE





Financement : ANR-10-IDEX-0004-02

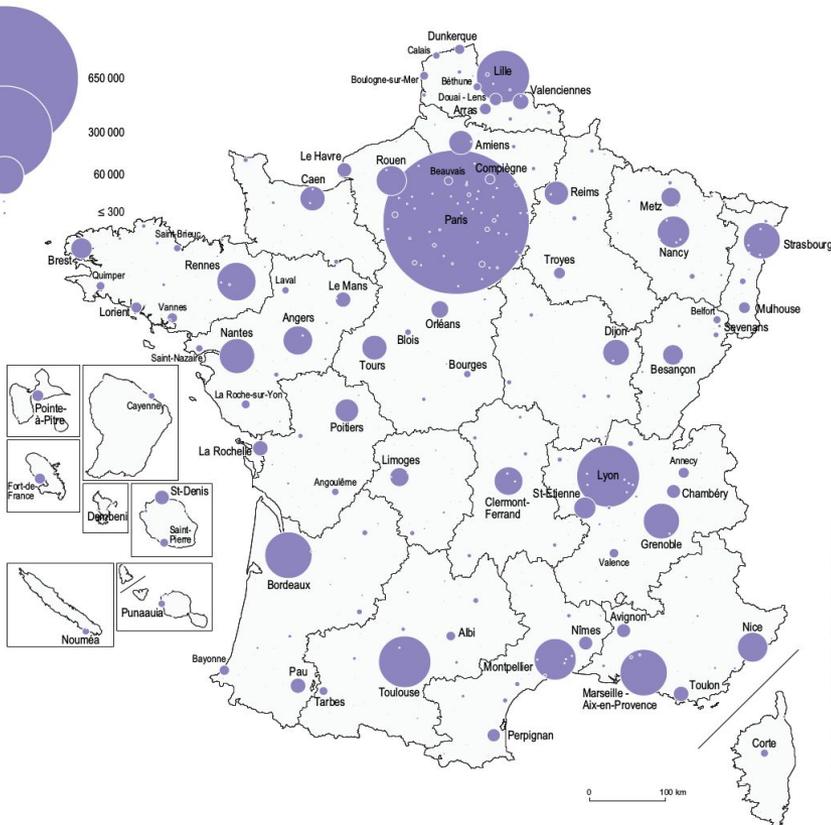


**55 millions d'euros
pour acquérir plus de**

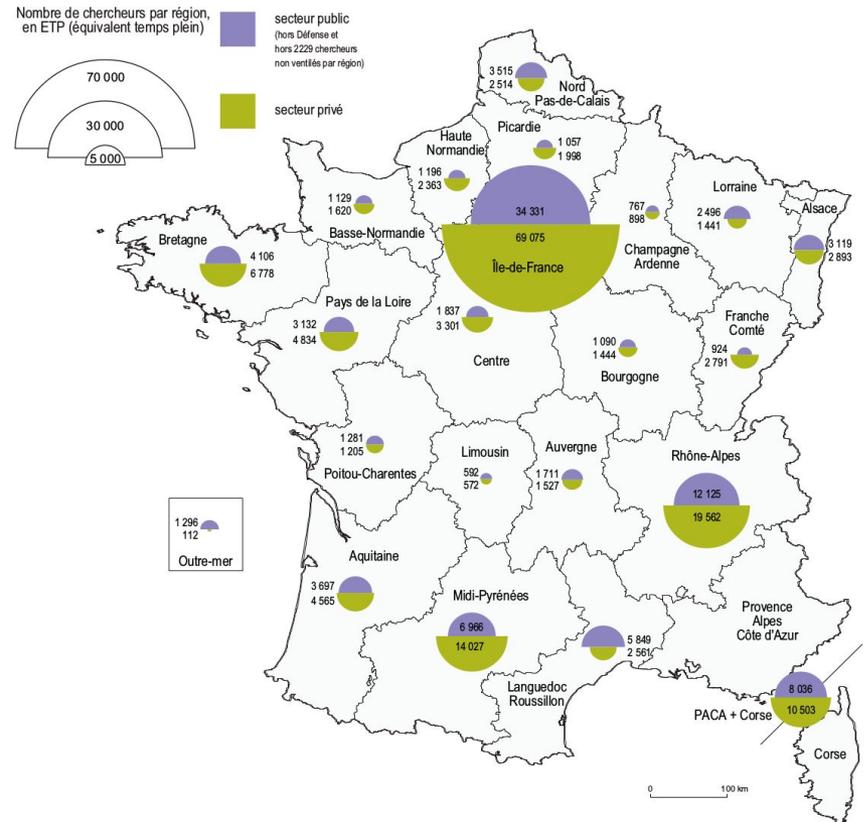
17 millions de documents numériques

Pour la recherche et l'enseignement supérieur français

Nombre d'étudiants inscrits par unité urbaine (Insee 2010)



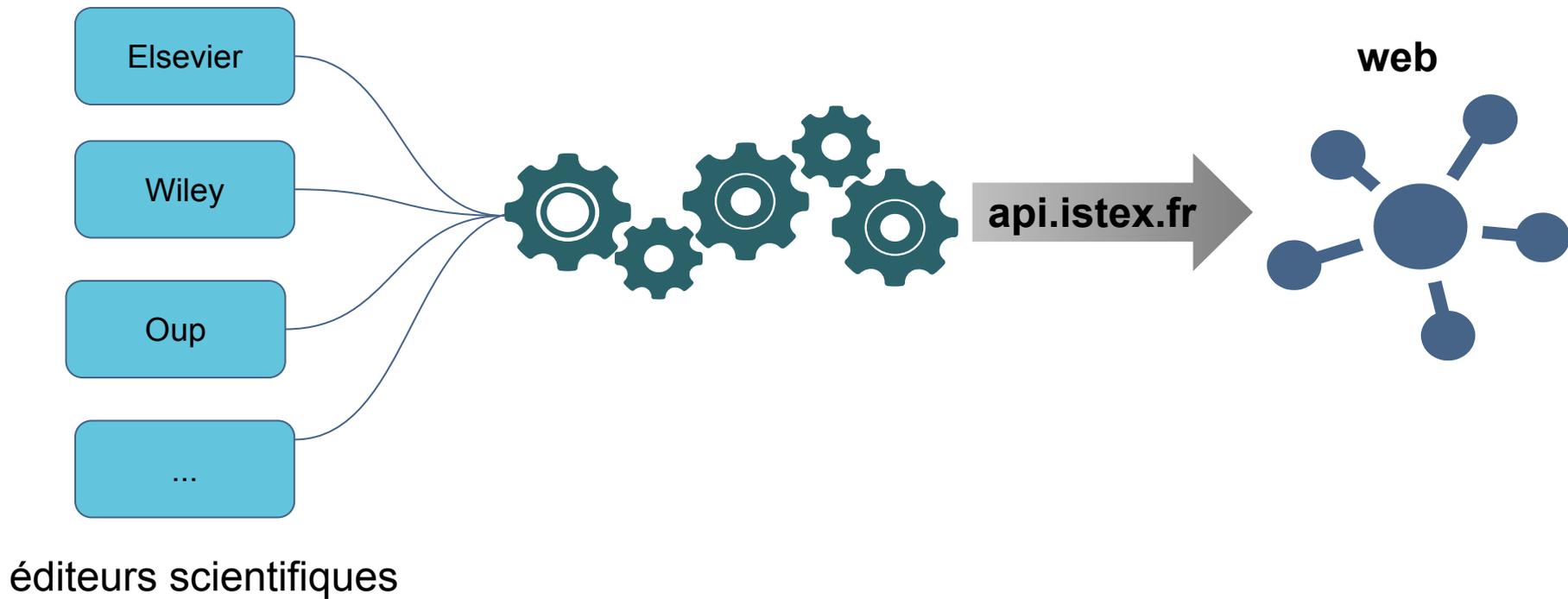
Nombre de chercheurs par région, en ETP (équivalent temps plein)



Source et réalisation : MENESR-DGESIP/DGRI-SIES

Source et réalisation : MENESR-DGESIP/DGRI-SIES

Avec une plateforme de curation & une API web



ISTEX : des objets numériques

891



3. Zur Elektrodynamik bewegter Körper; von A. Einstein.

Daß die Elektrodynamik Maxwells — wie dieselbe gegenwärtig aufgefaßt zu werden pflegt — in ihrer Anwendung auf bewegte Körper zu Asymmetrien führt, welche den Phänomenen nicht anzuhaften scheinen, ist bekannt. Man denke z. B. an die elektrodynamische Wechselwirkung zwischen einem Magneten und einem Leiter. Das beobachtbare Phänomen hängt hier nur ab von der Relativbewegung von Leiter und Magnet, während nach der üblichen Auffassung die beiden Fälle, daß der eine oder der andere dieser Körper der bewegte sei, streng voneinander zu trennen sind. Bewegt sich nämlich der Magnet und ruht der Leiter, so entsteht in der Umgebung des Magneten ein elektrisches Feld von gewissem Energiewerte, welches an den Orten, wo sich Teile des Leiters befinden, einen Strom erzeugt. Ruht aber der Magnet und bewegt sich der Leiter, so entsteht in der Umgebung des Magneten kein elektrisches Feld, dagegen im Leiter eine elektromotorische Kraft, welcher an sich keine Energie entspricht, die aber — Gleichheit der Relativbewegung bei den beiden ins Auge gefaßten Fällen vorausgesetzt — zu elektrischen Strömen von derselben Größe und demselben Verlaufe Veranlassung gibt, wie im ersten Falle die elektrischen Kräfte.

Beispiele ähnlicher Art, sowie die mißlungenen Versuche, eine Bewegung der Erde relativ zum „Lichtmedium“ zu konstatieren, führen zu der Vermutung, daß dem Begriffe der absoluten Ruhe nicht nur in der Mechanik, sondern auch in der Elektrodynamik keine Eigenschaften der Erscheinungen entsprechen, sondern daß vielmehr für alle Koordinatensysteme, für welche die mechanischen Gleichungen gelten, auch die gleichen elektrodynamischen und optischen Gesetze gelten, wie dies für die Größen erster Ordnung bereits erwiesen ist. Wir wollen diese Vermutung (deren Inhalt im folgenden „Prinzip der Relativität“ genannt werden wird) zur Voraussetzung erheben und außerdem die mit ihm nur scheinbar unverträgliche



```
<!-- Version 1.2 générée le 29-9-2016 -->
<?TEI xmlns:tei="http://www.tei-c.org/ns/1.0" xml:lang="de" type="article">
  <teiHeader>
    <fileDesc>
      <titleStm>
        <title level="a" type="main" xml:lang="de">Zur Elektrodynamik bewegter Körper</title>
      </titleStm>
      <publicationStm>
        <authority>ISTEX</authority>
        <publisher>WILEY-VCH Verlag</publisher>
        <pubPlace>Leipzig</pubPlace>
      <availability>
        <p>
          Copyright © 1905 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim
        </p>
      </availability>
      <date>1905</date>
    </publicationStm>
    <sourceDesc>
      <bibliStruct type="inbook">
        <analytic>
          <title level="a" type="main" xml:lang="de">Zur Elektrodynamik bewegter Körper</title>
          <author xml:id="author-1">
            <persName>
              <forename type="first">A.</forename>
              <surname>Einstein</surname>
            </persName>
            <affiliation>Bern</affiliation>
          </author>
        </analytic>
        <monogr>
          <title level="j">Annalen der Physik</title>
          <title level="j" type="abbrev">Ann. Phys.</title>
          <idno type="PISSN">0003-3804</idno>
          <idno type="eISSN">1521-3889</idno>
          <idno type="DOI">10.1002/(ISSN)1521-3889</idno>
        </monogr>
        <imprint>
          <publisher>WILEY-VCH Verlag</publisher>
          <pubPlace>Leipzig</pubPlace>
          <date type="published" when="1905"/>
          <biblScope unit="volume">322</biblScope>
          <biblScope unit="issue">10</biblScope>
          <biblScope unit="page" from="891">891</biblScope>
          <biblScope unit="page" to="921">921</biblScope>
        </imprint>
        <idno type="istex">21A763EE4C0CE534CED7E03A40E75865A527346</idno>
        <idno type="DOI">10.1002/andp.19053221004</idno>
        <idno type="ArticleID">ANDP19053221004</idno>
      </bibliStruct>
    </sourceDesc>
  </fileDesc>
  <profileDesc>
    <creation>
      <date>1905</date>
    </creation>
    <langUsage>
      <language ident="de">de</language>
    </langUsage>
    <textClass>
      <keywords scheme="Journal Subject">
        <list>
          <head>article-category</head>
          <item>
            <term>Article</term>
          </item>
        </list>
      </keywords>
    </textClass>
  </profileDesc>
```



composés de :

- ❖ **La notice bibliographique**
 - formats : Mods, TEI, éditeurs

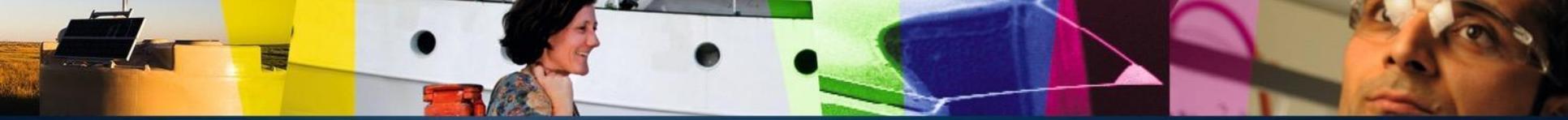
- ❖ **Le plein texte**
 - formats : PDF, TXT

- ❖ **Des annotations**
 - Entités nommées
 - Références citées
 - Indexation
 - Classification





Quid des standards du W3C ?



Postulats

- ◉ Volonté “politique”
- ◉ Pas de reformatage de l'existant
- ◉ Ne pas refaire un réservoir de documents



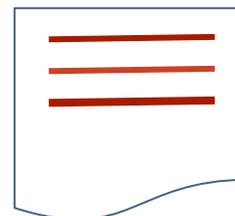
Lancement d'une expérimentation en janvier 2016
pour **tester** l'usage des normes du web sémantique

<http://lodex.inist.fr/>

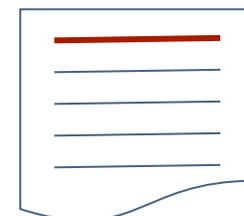
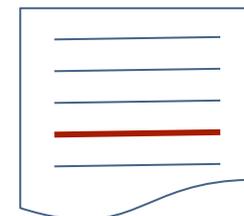
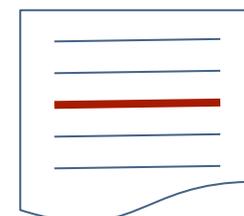
Avec quelles données ?

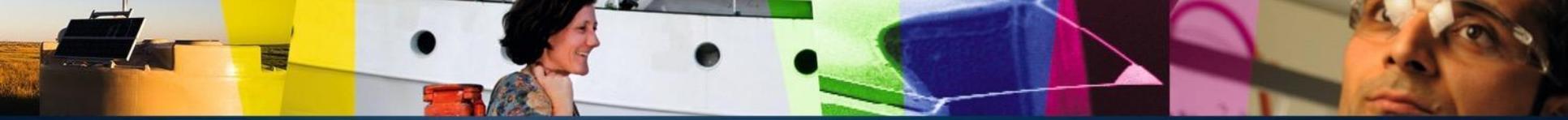
- Des jeux de documents à partir d'informations extraites automatiquement (entités nommés, catégories scientifiques, etc.)
- Récupérer toutes les informations induites et produites par les documentalistes (types de documents, regroupement des langues, etc.)

LODEX



ISTEX





Pour quel usage ?

Proposer une documentation du fonds ISTEK dynamique et interopérable

- Enrichir les services existants
- Améliorer l'accès et l'usage de l'API
- Valoriser toutes les informations non présentes dans les documents

Comment ?

Une petite équipe mixte associant informaticiens et documentalistes

- ◉ Proposer un guide méthodologique pour rendre les données “sémantiques”
- ◉ Développer un prototype pour les transformer et les visualiser

Avec le mode opératoire suivant:





Les 7 étapes du mode opératoire

1 - Choix de la licence



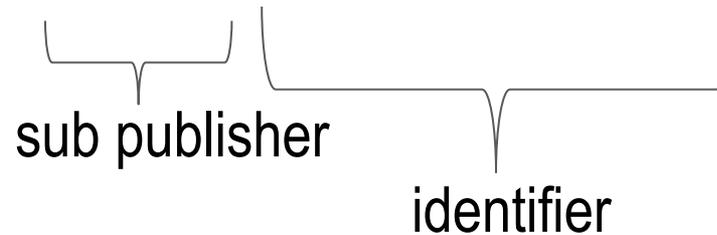
LICENCE OUVERTE

OPEN LICENCE

Licence Etalab pour les méta données,
Accès aux documents contrôlés

3 - Constitution des URI

ark:/67375/39D-S2GXG1TW-8



sub-publisher : Un identifiant pour chaque jeu de données

identifiant : un identifiant pour chaque “ressource” d’un jeu de données

4 - Analyse des données



Editeur
des données
(*équipe LODEX*)

Producteur
des données
(*documentalistes*)

2 objectifs :

- ◉ Minimiser la préparation des données sources
- ◉ Reformater les données sources à posteriori

5 - Enrichissement / Alignement des données

- Manuellement
par des experts sur des petits jeux de données
- Automatiquement
par des algorithmes spécifiques de “clustering”

Type	Entité Nom-mée	Statut	Population	Latitude	Longitude	Région	Pays	Continent	GéoName	Data BNF	Wikidata	Wikipedia	Geohack
Ville	Tokyo	Division administrative de premier ordre	12,445,327	N 35° 41' 22"	E 139° 41' 30"	Tokyo	Japon	Asie	http://sws.geo	http://data	https://www.w	https://en.w	
		Capitale	8,336,599						http://sws.geo		https://www.w		
Ville	San Diego	Division administrative de second ordre	1,307,402	N 32°42'55"	W 117°09'26"	Californie	USA	Amérique du Nord	http://sws.geo		https://www.w	https://en.w	
		Ville habitée	8014	N 10°20'10"	W 73°10'55"	Cesar	Colombie	Amérique du Sud	http://sws.geo			https://en.w	
		Division administrative de second ordre	13772	N 10° 13'57"	W 73°16'25"				http://sws.geo				
		Division administrative de second ordre		N 10° 15' 22"	W 67° 57' 14"	Carabobo	Vénézuela	http://sws.geo			https://en.w		
Union	UK	Independent political entity	62348447	N 54°45'30"	W 2°41'43"			Europe	http://sws.geo	http://data	https://www.w	https://en.w	https://tools.w
	United Kingdom												
	Europe	Union de pays							http://data	https://www.w			https://tools.w
Pays	Italy	Pays	60,340,328	N 42° 50' 0"	E 12° 50' 0"		Italie		http://sws.geo	http://data	https://www.w	https://en.w	https://tools.w
	United States	Pays	310,232,863	N 39° 45' 36"	W 98° 30' 0"		Etats-Unis	Amérique du Nord	http://sws.geo	http://data	https://www.w	https://en.w	https://tools.w
	USA												
Etat	IN			39° 56' 0" N	86° 13' 0" W					https://www.w		https://tools.w	
Continent	America			N 83°10'-S 56°	W 165°34'51'			Amérique		http://data	https://www.w		
Comté	Nottighamshire	Division administrative de second ordre	1,055,400	N 53°10'00"	W 1°00'00"		Angleterre	Europe	http://sws.geo	http://data		https://en.w	

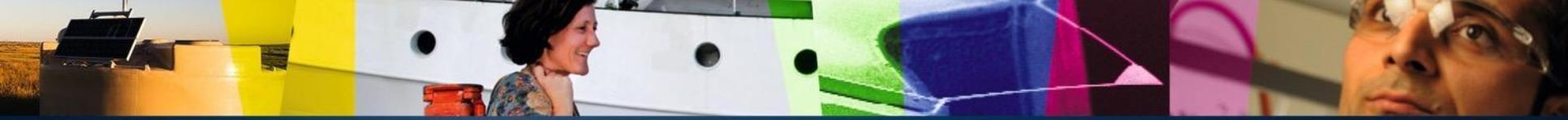
7 - Publication

Développement d'un prototype pour proposer un réseau de sites web autonomes et inter-connectés exposant chaque jeu de données

The screenshot shows the ISTEX website home page. At the top, there is a navigation bar with 'ISTEX' and various icons. Below it, a banner explains the LODEX experiment. The main content area features a 'Catégories INIST' section with a grid of category cards. On the right, there are two circular indicators: '127 RESSOURCES' and '13 PROPRIÉTÉS'. The bottom of the page has a 'Missing 'print:overview' in your config file' message and a grid of category cards.

The screenshot shows a detailed view of the 'NUCLEAR PHYSICS' category page. The page title is 'Catégories INIST' and 'NUCLEAR PHYSICS'. It includes a 'Description' section, a 'Resource' button, and a 'Share' button. Below, there are sections for 'Missing 'print:item.resource' in your config file', 'Properties', and 'Has top concept:'. The 'Properties' section shows 'ID: 1317205' and 'Title: NUCLEAR PHYSICS'. The 'Has top concept' section shows 'Catégories INIST' with '13 propriétés'. The 'Generic' section shows 'GENERAL ASPECTS' and 'Properties'. The 'Updated' section shows 'November 15, 2016'. At the bottom, it says 'No de document dans ISTEX : 1317205' and 'Dataset properties'.

The screenshot shows a detailed view of the 'Films' category page. The page title is 'ISTEX' and 'Films'. It includes a 'Description' section, a 'Topic' dropdown, and a 'Distribution per year' chart. Below, there are sections for 'Missing 'print:overview' in your config file', 'Overview', 'Tableau', 'Exports', and 'Ontology'. The 'Overview' section shows a grid of film cards with titles like 'Risky 1', 'Risky 2', 'Risky 3', 'Terminator', 'Terminator 2', 'Terminator 3', 'Terminator Salvation', and 'Total Recall'. The 'Director' section shows 'George Lucas' and 'James Cameron'.



Bilan

10 mois pour :

- ❖ tester la démarche sur des cas différents :
 - Sélection de 6 types d'information représentatifs
 - 2 équipes productrices différentes

- ❖ Tester l'usage et la réutilisation des données

- ❖ Identifier les blocages et des limites
 - pour monter en compétence
 - faire sauter les verrous un à un

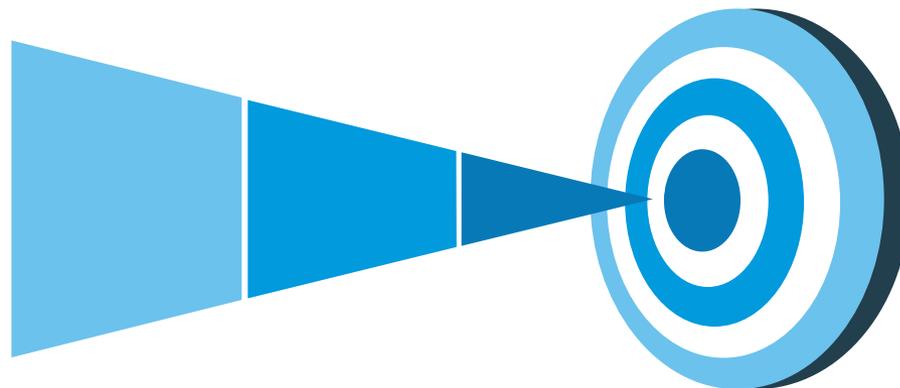


Validation du processus

Perspectives

On relance de 10 mois avec **un challenge** :

**Généraliser le processus
sur un maximum d'information
en rendant
les producteurs de données autonomes**



- *Ontologie spécifique*
- *Triplestore*
- ...



www.cnrs.fr

Merci

