

La TEI (dans) **CONDITOR**

■ référencement de la production scientifique française ■

Valérie Bonvallot
Merci à Christiane Stock
Inist-CNRS

ANF

La TEI pour l'IST : applications et perspectives en documentation

Mardi 3 décembre 2019

Un projet multipartenaire soutenu par le MESRI

CNRS (Dist, Inist, InSHS, SAP2S), Inra, Inria, IRD, Irstea

Univ. Angers, Bordeaux, Grenoble, Paris Dauphine, Paris Diderot, Sorbonne Université

Abes, Amue, CCSD, Hcéres (OST, DSI), Huma-Num, MESRI RNSR-ScanR

+ Univ. Limoges, Montpellier, Nancy, Nice, Strasbourg



**Recenser la production scientifique
de la recherche publique française**



Aux multiples bénéfices

Connaissance de la
production de l'ESR

Indicateurs de la science

Mieux connaître la production de son établissement

Qualité et complétude des
réservoirs ESR existants

Tâches de référencement

Limiter les multiples saisies au sein d'un établissement ou entre établissements

Qualité du RNSR

Référentiels communs

Cohérence SI

Accéder à toute source intégrée dans Conditor via un seul connecteur ...



Signalements sources

(Hal et archives institutionnelles, SI recherche ESR, bases de signalements...)

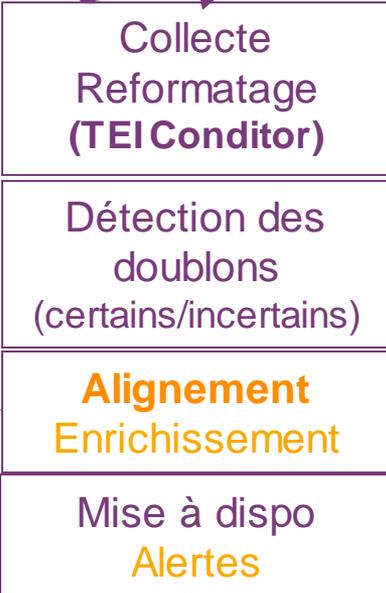
Hal, PubMed, CrossRef, thèses du Sudoc (années 2014 à 2017)

au fil de l'eau

Service d'alignement partenaire



Outillage



Référentiels RNSR...

Signalements sources enrichis
Signalements complémentaires



API



Réservoir de signalements Conditor



Organisation multipartenaire

Préfiguration du réseau métier

IHM : validation : Cornelius, suivi : Kibana

+ [Wiki](#)

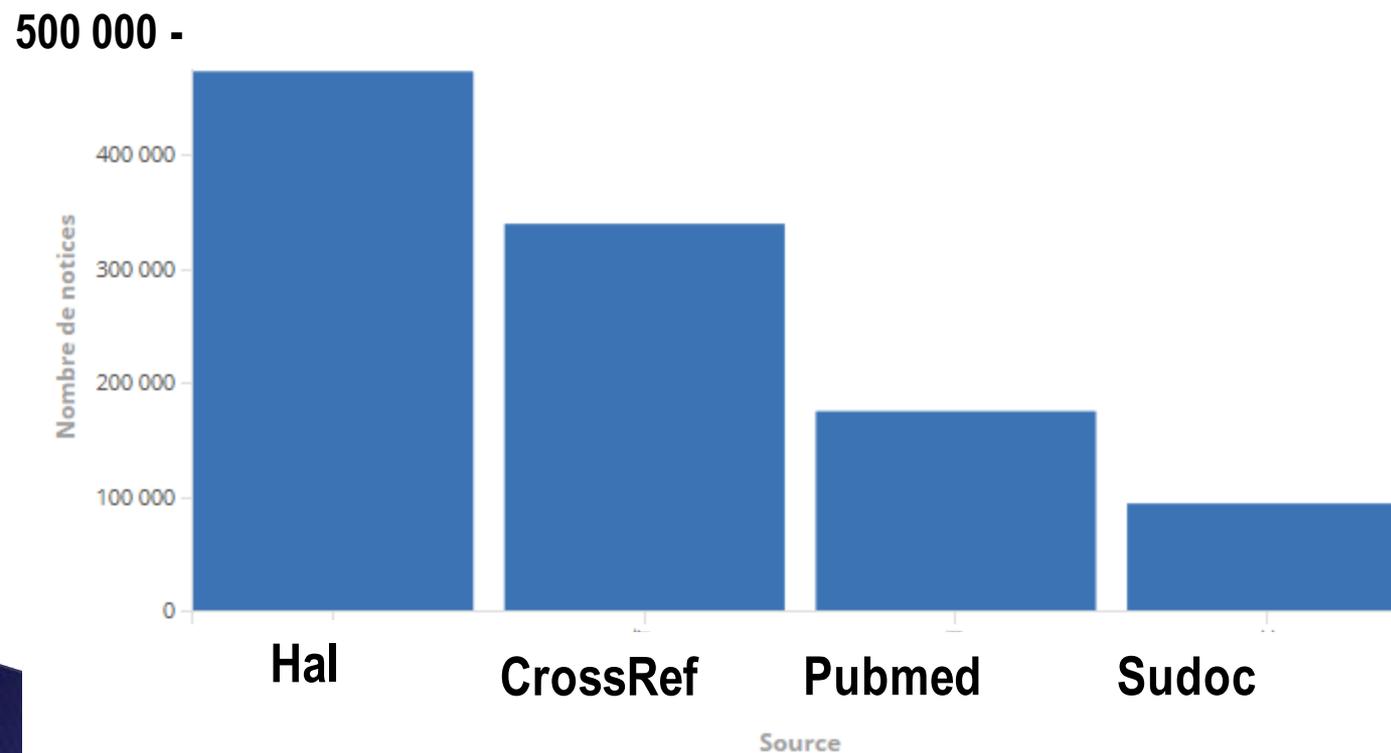


A faire / En cours
Hors plateforme Conditor

Aucune correction de métadonnées provenant de signalements sources

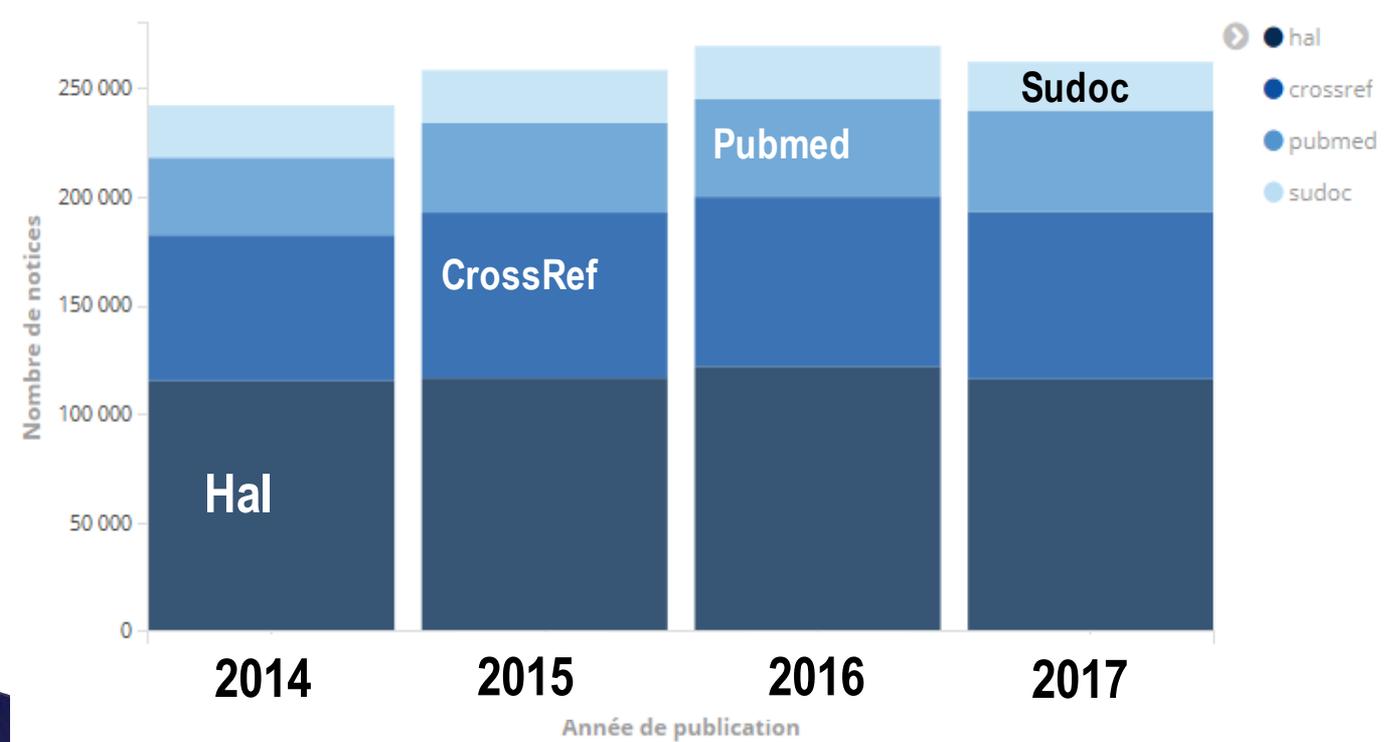
Nb de notices	Nb de doublons certains	Nb de notices uniques	Nb de doublons incertains
1 084 798 Nombre total de notices	523 886 Nombre de doublons certains	795 940 Nombre de publications uniques	122 736 Nombre de notices à valider

Origine du signalement

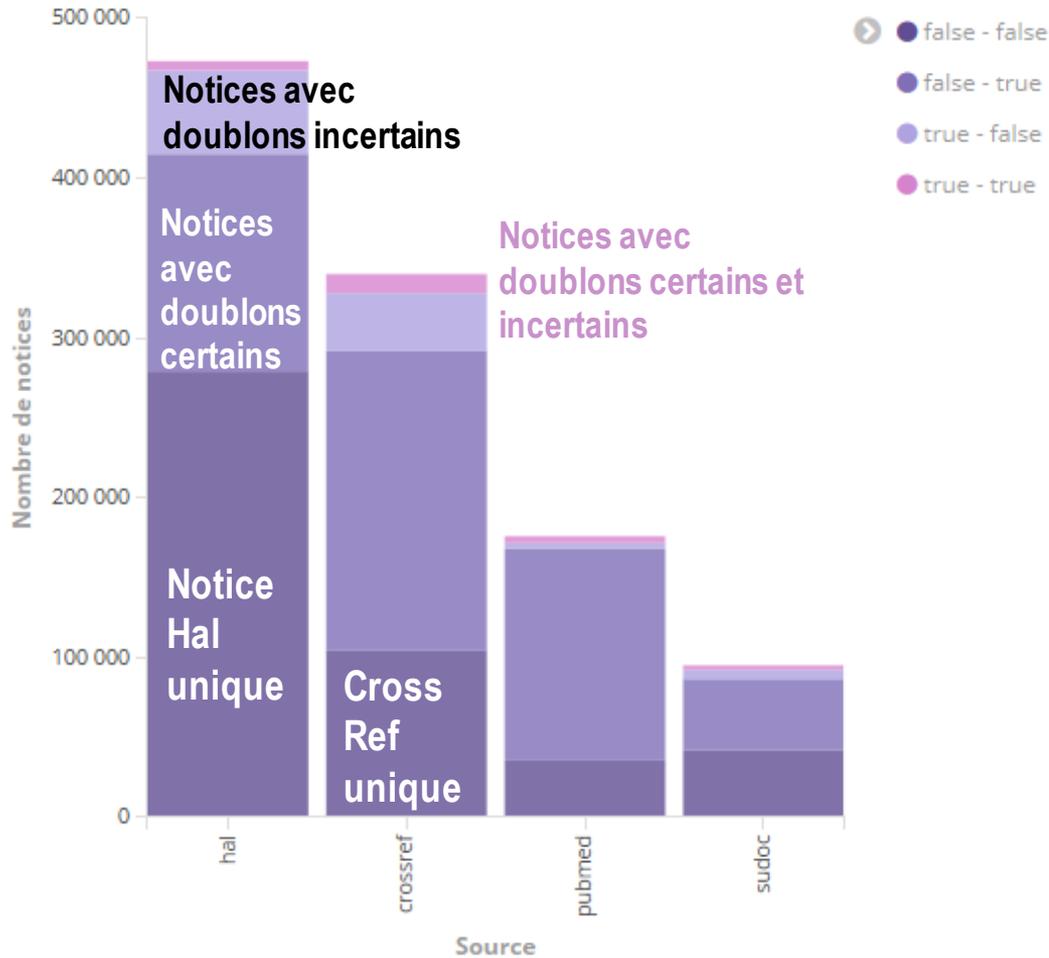


Nombre t	Nb de notices	Nb de doublons certains	or	Nb de notices uniques	Nb de doublons incertains
	1 084 798	523 886		795 940	122 736
	Nombre total de notices	Nombre de doublons certains		Nombre de publications uniques	Nombre de notices à valider

Année de publication (papier 2014-2017) et source



Répartition des doublons par base



La TEI (dans) CONDITOR



CONDITOR
référencement de la production scientifique française



XML1

Programmes de reformatage « maison »



XML2



XML3

Feuilles de styles



XML4

```
<TEI>
<text>
  <body>
    <listBib>
      <bibFull>
        <sourceDesc/>
        <profileDesc/>
      </bibFull>
    </listBib>
  </body>
</text>
</TEI>
```

```
<TEI>
<text>
  <body>
    <listBib>
      <biblFull>
        <sourceDesc/>
        <profileDesc/>
      </biblFull>
    </listBib>
  </body>
</text>
</TEI>
```

```
<TEI>...
  <sourceDesc>
    <biblStruct>
      <analytic>
        <monographic>
          <idno>
```

```
<TEI>...
  <sourceDesc>
    <biblStruct>
      <analytic>
        <title>
        <author>
          <persName>
          <affiliation>
          ....
```

```
<TEI>
<text>
  <body>
    <listBibl>
      <biblFull>
        <sourceDesc/>
        <profileDesc/>
      </biblFull>
    </listBibl>
  </body>
</text>
</TEI>
```

```
<TEI>...
  <sourceDesc>
    <biblStruct>
      <analytic>
        <monogr>
          <idno>
```

```
<TEI>...<sourceDesc><biblStruct>
  <analytic>
    <monogr>
      <idno type="issn">
      <idno type="eissn">
      <title level="j">
      <imprint>
        <publisher>
          <biblScope unit="volume">
          <biblScope unit="issue">
          <biblScope unit="pp">
          <date type="datePub">
          <date type="dateEpub">
```

```
<TEI>
<text>
  <body>
    <listBibl>
      <biblFull>
        <sourceDesc/>
        <profileDesc/>
      </biblFull>
    </listBibl>
  </body>
</text>
</TEI>
```

```
<TEI>...
<sourceDesc>
  <biblStruct>
    <analytic>
    <monogr>
      <idno>
```

```
<TEI>...
<sourceDesc>
  <biblStruct>
    <analytic>
    <monogr>
      <idno>
        <idno type="halId">
        <idno type="doi">
          ...
    </biblStruct>
  </sourceDesc>
```

```
<TEI>
<text>
  <body>
    <listBibl>
      <biblFull>
        <sourceDesc/>
        <profileDesc/>
      </biblFull>
    </listBibl>
  </body>
</text>
</TEI>
```

```
<TEI>...
  <profileDesc>
    <textClass>
      <keywords scheme="rameau">
      <keywords scheme="fmesh">
      <classCode scheme="thesisDomain">
      <classCode scheme="typology">
    </textClass>
    <abstract xml:lang=""></abstract>
  </profileDesc></biblFull></listBibl>
</body></text></TEI>
```

```
<?xml version="1.0"?>
```

```
<crossref xmlns="http://www.crossref.org/xschema/1.1" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  <journal>
    <journal_metadata>
      <full_title>SHS Web of Conferences</full_title>
      <abbrev_title>SHS Web of Conferences</abbrev_title>
      <issn media_type="electronic">2261-2424</issn>
    </journal_metadata>
    <journal_issue>
      <publication_date>
        <year>2014</year>
      </publication_date>
      <journal_volume>
        <volume>8</volume>
      </journal_volume>
    </journal_issue>
    <journal_article publication_type="full_text" language="fr">
      <titles>
        <title>Narrations écrites et orthographe chez les enfants et adolescents dysphasiques :
      </titles>
      <contributors>
        <person_name sequence="first" contributor_role="author">
          <given_name>Lucie</given_name>
          <surname>Broc</surname>
        </person_name>
        <person_name sequence="additional" contributor_role="author">
          <given_name>Josie</given_name>
          <surname>Bernicot</surname>
        </person_name>
        ...
      </contributors>
```





```

<sourceDesc>
  <bibliStruct>
    <analytic>
      <title>
        Narrations écrites et orthographe chez les enfants et adolescents dysphasiques : Pourquoi les erreurs morphologiques augmentent-elles avec l'âge ?
      </title>
      <author role="aut">
        <persName>
          <forename type="first">Lucie</forename>
          <surname>Broc</surname>
        </persName>
      </author>
      <author role="aut">
        <persName>
          <forename type="first">Josie</forename>
          <surname>Bernicot</surname>
        </persName>
      </author>
      ...
    </analytic>
  </bibliStruct>
</sourceDesc>

```

```

<monogr>
  <idno type="eissn">2261-2424</idno>
  <title level="j">SHS Web of Conferences</title>
  <title level="j" type="abbreviso">SHS Web of Conferences</title>
  <imprint>
    <date type="datePub">2014</date>
    <date type="dateEpub">2014</date>
    <biblScope unit="volume">8</biblScope>
    <biblScope unit="pp">1395-1405</biblScope>
  </imprint>
</monogr>
<idno type="doi">10.1051/shsconf/20140801167</idno>
<idno type="pii">shsconf_cmlf14_01167</idno>
</bibliStruct>
</sourceDesc>
<profileDesc>
  <textClass>
    <classCode scheme="typology">Journal article</classCode>
  </textClass>
</profileDesc>

```

Le



(dans)

CONDITOR
référencement de la production scientifique française

JSON : *JavaScript Object Notation* - format d'échange de données

Extraction de champs pour les traitements de repérage des doublons et les alignements avec le RNSR

Champs interrogeables :
elasticSearch



Kibana



API



Cornelius

Applicatifs
partenaires

La TEI complète dans un champ
Des champs de la TEI

- Identifiants de l'article
- Titre de l'article
- Auteurs et affiliations + identifiants
- Titre du périodique (+issn, + eissn)
- Volumaison (volume, numéro, page)

Des nouveaux champs

- Type de document Conditor
- Repérage des doublons certains
- Repérage des doublons incertains
- ...

[Tous les champs json](#)

Le



(dans)

CONDITOR

référencement de la production scientifique française

```

"creationDate": "2019-06-28 05:44:49",
"title": {
  "default": "Narrations écrites et orthographe chez les enfants et adolescents dysphasiques : Pourquoi les erreurs morphologiques augmentent-elles avec l'âge ?",
  "fr": "",
  "en": "",
  "monography": "",
  "journal": "SHS Web of Conferences",
  "meeting": ""
},
"first3AuthorNames": "Broc Lucie Bernicot Josie Olive Thierry",
"authorNames": "Broc Lucie Bernicot Josie Olive Thierry Favart Monik Reilly Judy Quémart Pauline Uzé Joel",
"first3AuthorNamesWithInitials": "Broc L Bernicot J Olive T",
"authors": [
  {
    "forename": "Lucie",
    "isni": [],
    "surname": "Broc",
    "researcherId": "",
    "halAuthorId": "",
    "affiliations": [],
    "orcid": [],
    "idHal": "",
    "idRef": [],
    "viaf": []
  },
  {
    "forename": "Josie",
    "isni": [],
    "surname": "Bernicot",
    "researcherId": "",
    "halAuthorId": "",
    "affiliations": [],
    "orcid": [],
    "idHal": "",
    "idRef": [],
    "viaf": []
  }
]
}

```

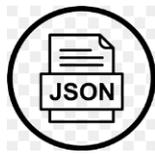
```

<?xml version="1.0" encoding="UTF-8"?>
<TEI xmlns="http://www.tei-c.org/ns/1.0"><text><body><listBibl><biblFull><titleStmt><title>Narrations écrites et orthographe chez les enfants et adolescents dysphasiques : Pourquoi les erreurs morphologiques augmentent-elles avec l'âge ?</title></titleStmt><editionStmt><edition><date type="whenDownloaded">2019-02-11</date><date type="whenCreated">2019-05-24</date></edition></editionStmt><publicationStmt><distributor>Conditor</distributor></publicationStmt><sourceDesc><biblStruct><analytic><title>Narrations écrites et orthographe chez les enfants et adolescents dysphasiques : Pourquoi les erreurs morphologiques augmentent-elles avec l'âge ?</title><author role="aut"><persName><forename type="first">Lucie</forename><surname>Broc</surname></persName></author><author role="aut"><persName><forename type="first">Josie</forename><surname>Bernicot</surname></persName></author><author role="aut"><persName><forename type="first">Thierry</forename><surname>Olive</surname></persName></author><author role="aut"><persName><forename type="first">Monik</forename><surname>Favart</surname></persName></author><author role="aut"><persName><forename type="first">Judy</forename><surname>Reilly</surname></persName></author><author role="aut"><persName><forename type="first">Pauline</forename><surname>Quémart</surname></persName></author><author role="aut"><persName><forename type="first">Joel</forename><surname>Uzé</surname></persName></author><author role="edt"><persName><forename type="first">Anne tte</forename><surname>Gerstenberg</surname></persName></author><author role="edt"><persName><forename type="first">Franck</forename><surname>Neveu</surname></persName></author><author role="edt"><persName><forename type="first">Judith</forename><surname>Meinschaefer</surname></persName></author><author role="edt"><persName><forename type="first">Linda</forename><surname>Hriba</surname></persName></author><author role="edt"><persName><forename type="first">P eter</forename><surname>Blumenthal</surname></persName></author><author role="edt"><persName><forename type="first">Sophie</forename><surname>Prévost</surname></persName></author></analytic><monogr><idno type="eissn">2261-2424</idno><title level="j">SHS Web of Conferences</title><title level="j" type="abbreviso">SHS Web of Conferences</title><imprint><date type="datePub">2014</date><date type="dateEpub">2014</date><biblScope unit="volume">8</biblScope><biblScope unit="pp">1395-1405</biblScope></imprint></monogr><idno type="doi">10.1051/shsconf/20140801167</idno><idno type="pii">shsconf_cmlf14_01167</idno></biblStruct></sourceDesc><profileDesc><textClass><classCode scheme="typology">Journal article</classCode></textClass></profileDesc></biblFull></listBibl></body></text></TEI>

```

Notice en TEI : champ

Le



(dans)

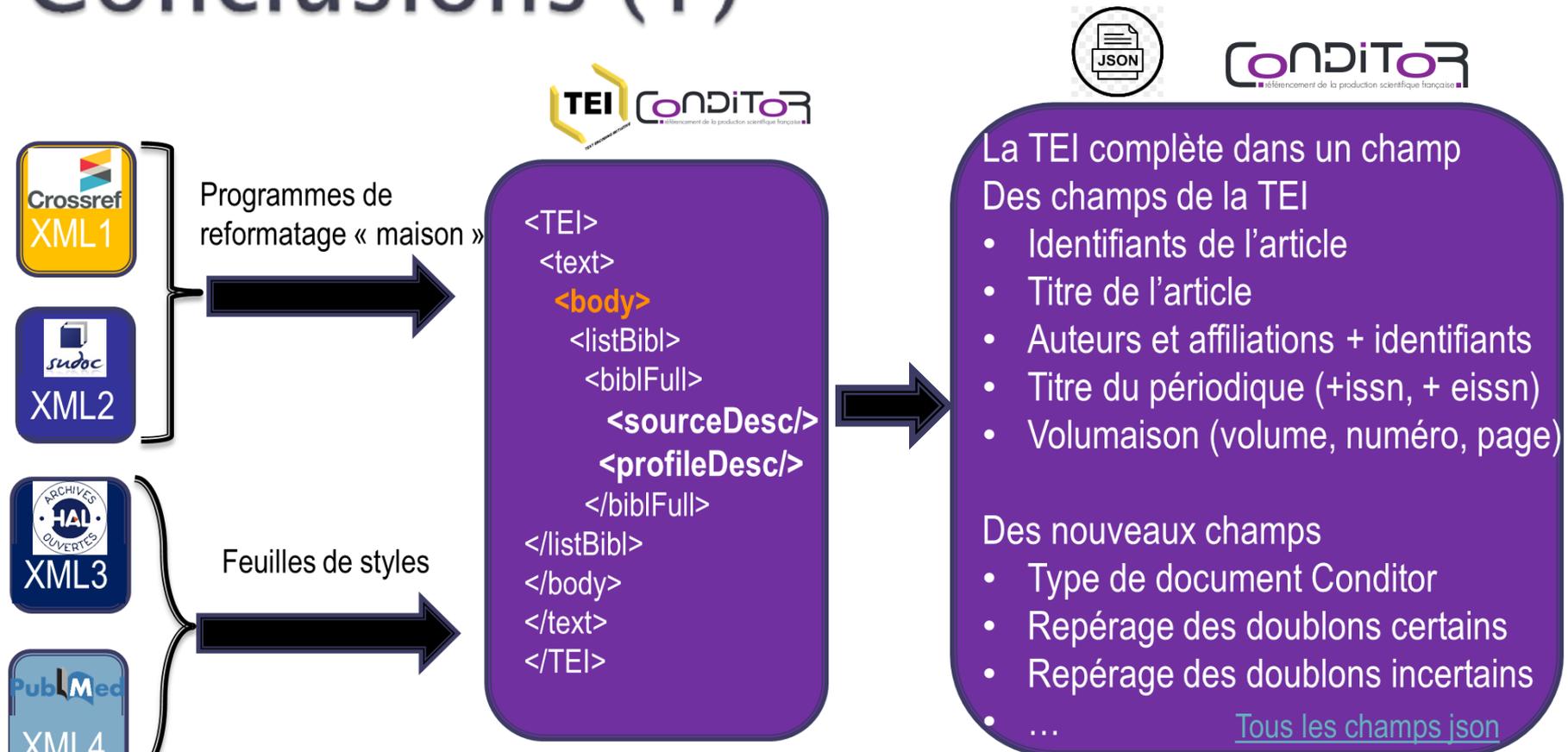
CONDITOR

référencement de la production scientifique française

```
"duplicates": [
  {
    "sessionName": "HAL_2019-06-27_2014_WITH_PUBDATE",
    "rules": [
      "Article : 2Collation:Tic+xissn+volume+page+annee",
      "Article : 2Collation:Tic+Ti_source+volume+page+annee"
    ],
    "source": "hal",
    "idConditor": "2USgHC_EDxKtvB91eH2M7uMFy"
  }
],
```

```
"isNearDuplicate": false,
"nearDuplicates": [],
"nearDuplicatesDetectedBySimilarity": [
  {
    "similarityRate": 1,
    "source": "hal",
    "type": "Article",
    "idConditor": "2USgHC_EDxKtvB91eH2M7uMFy"
  }
]
```

Conclusions (1)



Champs interrogeables :
elasticSearch – Kibana - API



Applicatifs partenaires

Conclusions (2)

- ▶ **Description** d'un objet qui n'est pas du texte intégral mais une **notice**
- ▶ Objectif de **reformatage** pour obtenir un **format unique**
 - Traitements
 - Interopérabilité
- ▶ **TEI** pas seule dans cet écosystème : **json**
- ▶ **Enrichissements** avec les codes RNSR : vers l'implémentation de **standOff ?**

Merci de votre attention



valerie.bonvallot@inist.fr

contact.conditor@services.cnrs.fr

Wiki <https://wiki.conditor.fr>