

# OUTIL D'ÉVALUATION DE L'USAGE DES RESSOURCES numériques



**couperin.org**  
Consortium Unifié des Établissements Universitaires et  
de Recherche pour l'Accès aux Publications Numériques



**UNIVERSITÉ  
DE LORRAINE**



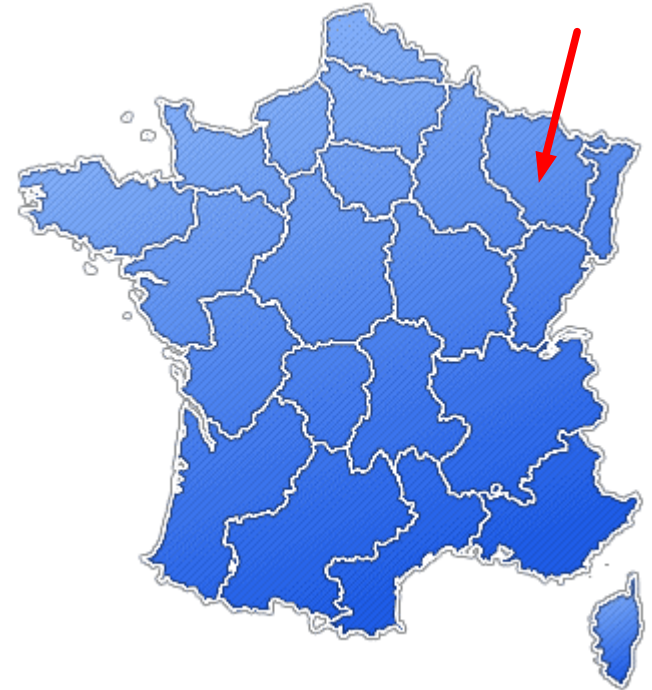
<http://analogist.couperin.org>

<http://ezparse.couperin.org>

# Qui sommes nous ?



- Dominique Lechaudel  
<[dominique.lechaudel@inist.fr](mailto:dominique.lechaudel@inist.fr)>
- Stéphane Gully  
<[stephane.gully@inist.fr](mailto:stephane.gully@inist.fr)>



# Plan



- Historique
- Démarche projet
- Présentation d'ezPAARSE/AnalogIST
- Coeur d'ezPAARSE : les parseurs
- Exploitation des résultats
- Écosystème d'ezPAARSE

# Rappel du besoin



- Cf : Quantifier les accès à la documentation numérique payante : un panorama des enjeux (Monique Joly, Grégory Colcanap, Thomas Porquet)

# Plan



- **Historique**
- Démarche projet
- Présentation d'ezPAARSE/AnalogIST
- Coeur d'ezPAARSE : les parseurs
- Exploitation des résultats
- Écosystème d'ezPAARSE

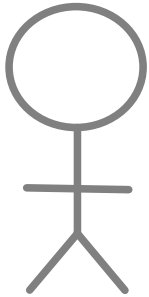
# Portails INIST



# Portails INIST



Chercheur



Des revues  
lui sont présentées.

Éditeur

SciVerse ScienceDirect Hub ScienceDirect Scopus ScITopics Applications

Home | Browse | Search | My settings | My alerts

Articles All fields Author Volume Issue

Journal/Book title

Browse 10,549,005 Articles

Browse by title

ABCDEFGHIJKLMNOPQRSTUVWXYZ

Browse by subject

- Physical Sciences and Engineering
  - Chemical Engineering
  - Chemistry
  - Computer Science
  - Earth and Planetary sciences
  - Energy
  - Engineering
  - Materials Science
  - Mathematics
  - Physics and Astronomy
- Life Sciences
  - Agricultural and Biological Sciences
  - Biochemistry, Genetics and Molecular Biology
  - Environmental Science
  - Immunology and Microbiology
  - Neuroscience
- Health Sciences
  - Medicine and Dentistry
  - Nursing and Health Professions
  - Pharmacology, Toxicology and Pharmaceutical Science

Updated on November 1st

Top 25 Hottest

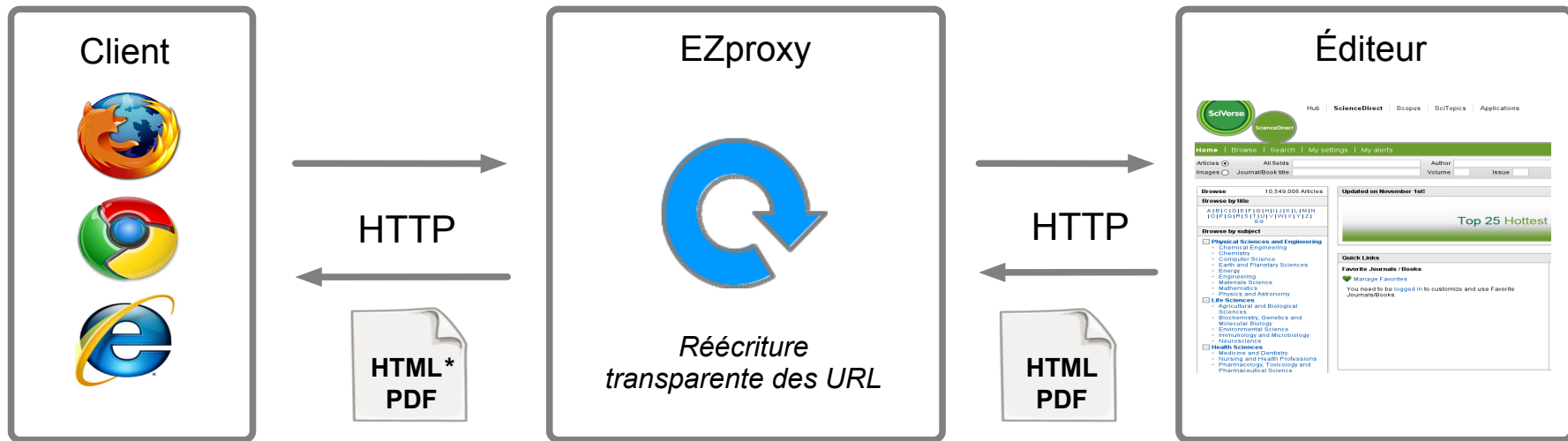
Quick Links

Favourite Journals / Books

Manage Favourites

You need to be logged in to customize and use Favorite Journals/Books.

# Fonctionnement d'EZProxy





# NUM\_STAT (2009-2012)



## Constats

- Les indicateurs produits à partir de NUM\_STAT font envie à tous les gestionnaires de ressources électroniques lors des communications
- La solution **statistiques locales** est viable
- Il n'existe pas de logiciel qui réponde complètement à ce besoin
- La maintenance de l'applicatif au niveau INIST-CNRS est très lourde : souvent dans le **rouge**

# Démarche projet



- Historique
- **Démarche projet**
- Présentation d'ezpaarse/analogist
- Coeur d'ezPAARSE : les parseurs
- Exploitation des résultats
- Écosystème d'ezPAARSE

# Initiation du projet ezPAARSE / AnalogIST



couperin.org  
Consortium Universitaire de Publications Numériques



Mai 2013

Lundi	Mardi	Mercredi	Jeudi	Vendredi	Samedi	Dimanche
		1	2	3	4	5
6	7	8	9	10	11	12
13	14	Mai 2012		18	19	
20	21	22	23	24	25	26
27	28	29	30	31		



## Vision du produit

"ezPAARSE permet l'analyse, l'enrichissement et l'exploitation des logs d'accès collectés localement aux ressources électroniques proposées par les plateformes Web des éditeurs.

Les utilisateurs d'ezPAARSE sont les services de documentation de l'enseignement supérieur et de la recherche (ESR). Les premiers utilisateurs seront l'INIST-CNRS et la direction de la documentation et de l'édition de l'Université de Lorraine en partenariat avec COUPERIN.

ezPAARSE répond aux besoins suivants : ...."

## Liste des fonctionnalités

Product Backlog d'ezPAARSE (EPIC)

1. Instance nationale publique ezPAARSE
2. Organisation fonctionnelle du réseau ezPAARSE
3. Instance locale ezPAARSE
4. Recommandations COUNTER
5. Interopérabilité avec MESURE
6. Reconnaissance du format de la ligne de log ....



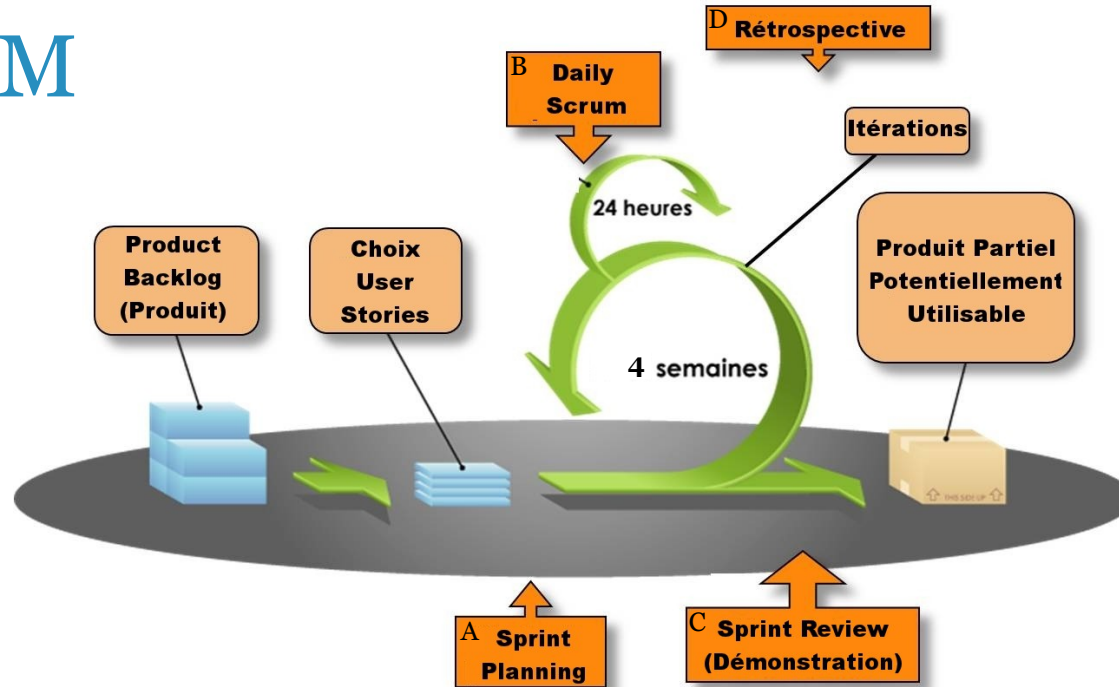
# L'équipe projet



# La méthode agile de développement



## SCRUM



# Les versions 0.x d'ezPAARSE



En méthode agile, à chaque sprint peut correspondre une version du produit,

En voici la liste qui a conduit à la version 1.0 d'ezPAARSE :

- 0.1 du 12/11/2012 au 06/12/2012 - Produit Minimum Viable
- 0.2 du 10/12/2012 au 20/12/2012 - Installation générique
- 0.3 du 07/01/2013 au 31/01/2013 - Extension du domaine des parseurs
- 0.4 du 04/02/2013 au 21/02/2013 - Extension des cas d'usage
- 0.5 du 04/03/2013 au 27/03/2013 - Utilisabilité d'ezPAARSE
- 0.6 du 02/04/2013 au 18/04/2013 - Consolidation des reconnaissances de plateforme
- 0.7 du 21/04/2013 au 05/06/2013 - Consolidation du coeur
- 0.8 du 10/06/2013 au 27/06/2013 - Qualification des utilisateurs
- 1.0 du 01/07/2013 au 08/08/2013 - Dédoublonnage COUNTER des EC

# Les versions 1.x d'ezPAARSE



...et tout est planifié

En voici la liste qui conduira à la version 2.0 d'ezPAARSE :

- 1.1 du 09/09/2013 au 10/10/2013 - MVP de synchronisation des PKB
- 1.2 du 14/10/2013 au 14/11/2013 - Sécurisation
- 1.3 du 18/11/2013 au 19/12/2013 - Multilinguisme
- 1.4 du 23/12/2013 au 23/01/2014 - Export COUNTER
- 1.5 du 27/01/2014 au 20/02/2014 - Intitulé à déterminer
- 1.6 du 24/02/2014 au 27/03/2014 - Intitulé à déterminer
- 1.7 du 31/03/2014 au 24/04/2014 - Intitulé à déterminer
- 1.8 du 28/04/2014 au 22/05/2014 - Intitulé à déterminer
- 1.9 du 26/05/2014 au 26/06/2014 - Intitulé à déterminer

# Produit minimal viable



```
analogist@stats: ~/ezparse/usr/share/ezparse
{"domain": "www.sciencedirect.com",
 "issn": "0092-8674"
},
{"ebef20411d29df128b106e3373aa685e": {
 "host": "b3a25f74a204d449cdbd9b81bdb28123",
 "login": "BVIUMR8161",
 "date": "2012-11-30T23:56:12+01:00",
 "url": "http://www.sciencedirect.com:80/science/article/pii/S0006291X12003440",
 "httpCode": "200",
 "size": "122403",
 "domain": "www.sciencedirect.com",
 "issn": "0006-291X"
},
{"92127bf36b259ceba5b8d16c08406c7e": {
 "host": "0179c5119b0e16f30ff472151c9ca340",
 "login": "BVIUMR7261",
 "date": "2012-12-01T00:00:33+01:00",
 "url": "http://www.sciencedirect.com:80/science/article/pii/S0965174804001821",
 "httpCode": "200",
 "size": "147679",
 "domain": "www.sciencedirect.com",
 "issn": "0965-1748"
},
{"b85f937bf43e790bb0539e4a213b3183": {
 "host": "d61a4e30f729943a29684bd3f7005bc3",
 "login": "BVIUMR7355",
 "date": "2012-12-01T00:02:53+01:00",
 "url": "http://www.sciencedirect.com:80/science/article/pii/S0024320512003268",
 "httpCode": "200",
 "size": "122487",
 "domain": "www.sciencedirect.com",
 "issn": "0024-3205"
}}
* Connection #0 to host 127.0.0.1 left intact
* Closing connection #0
analogist@stats:~/ezparse/usr/share/ezparse$
```



# Version en cours



- Contactez-nous par e-mail
- Visitez notre plateforme AnalogST
- Consultez la documentation d'ezPAARSE
- Suivez-nous sur Facebook
- Suivez-nous sur Twitter
- Suivez-nous sur Google+



- Contactez-nous par e-mail
- Visitez notre plateforme AnalogST
- Consultez la documentation d'ezPAARSE
- Suivez-nous sur Facebook
- Suivez-nous sur Twitter
- Suivez-nous sur Google+



▼ Traiter vos fichiers de log

Sélect. fichiers | Aucun fichier choisi

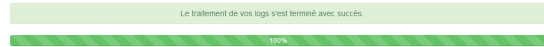
Glisser ici

► Options avancées

Traiter



- Contactez-nous par e-mail
- Visitez notre plateforme AnalogST
- Consultez la documentation d'ezPAARSE
- Suivez-nous sur Facebook
- Suivez-nous sur Twitter
- Suivez-nous sur Google+



Télécharger le résultat en csv

Fichiers traités

État du traitement

fede.bibliovie.exproxy.2013.05.29.log

Lignes en entrée	635721
ECs générés	8429
Plateformes reconnues	6
PDF	1854
HTML	3196
Taux de filtrage	97.66 %
Durée du traitement	15 m 15 s
Vitesse du traitement	695 lignes/s

Vos logs sont en cours de traitement mais vous pouvez déjà commencer à télécharger le résultat.

27%

Télécharger le résultat en csv

Annuler

Fichiers traités

fede.bibliovie.exproxy.2013.05.29.log

État du traitement

Lignes en entrée	173864
ECs générés	2158
Plateformes reconnues	6
PDF	397
HTML	811

# Présentation d'ezPAARSE/AnalogIST



- Historique
- Démarche projet
- **Présentation d'ezPAARSE/AnalogIST**
- Coeur d'ezPAARSE : les parseurs
- Exploitation des résultats
- Écosystème d'ezPAARSE

# ezPAARSE et AnalogIST



- **ezPAARSE** : le logiciel

(ez : easy / PAARSE : Progiciel d'Analyse des Accès aux Ressources Electroniques)

- en version locale : in situ
- en version globale : SaaS

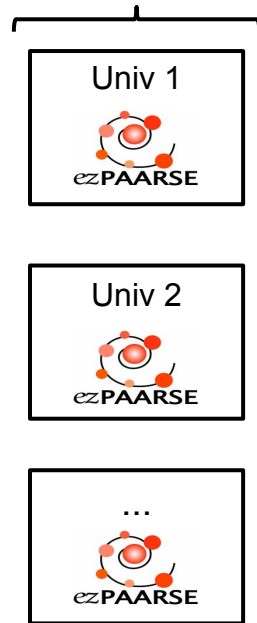
- **AnalogIST** : le portail collaboratif

(Analyse des Logs de l'IST) national, lieu de publication des analyses de plateforme, synchronisable avec les instances locales, installées au sein des institutions intéressées.

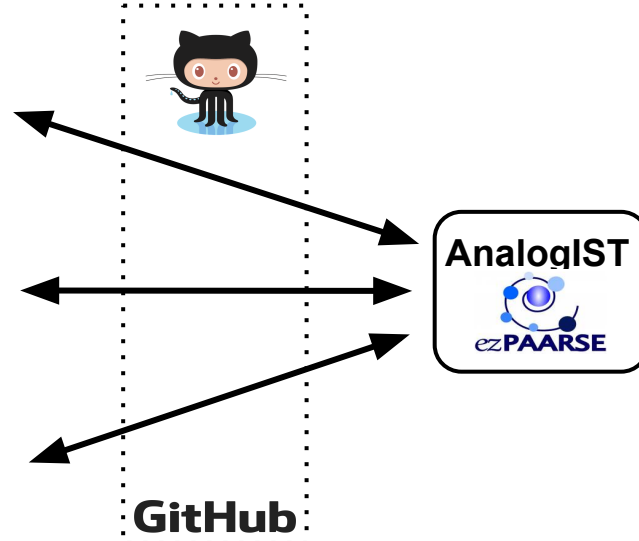
# AnalogIST vs ezPAARSE



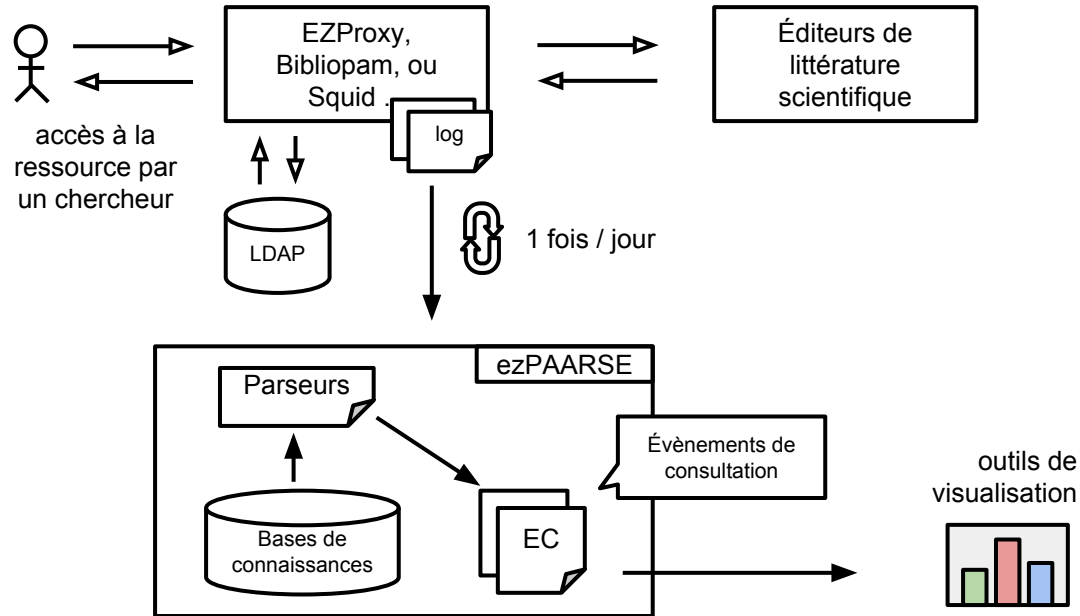
instances locales



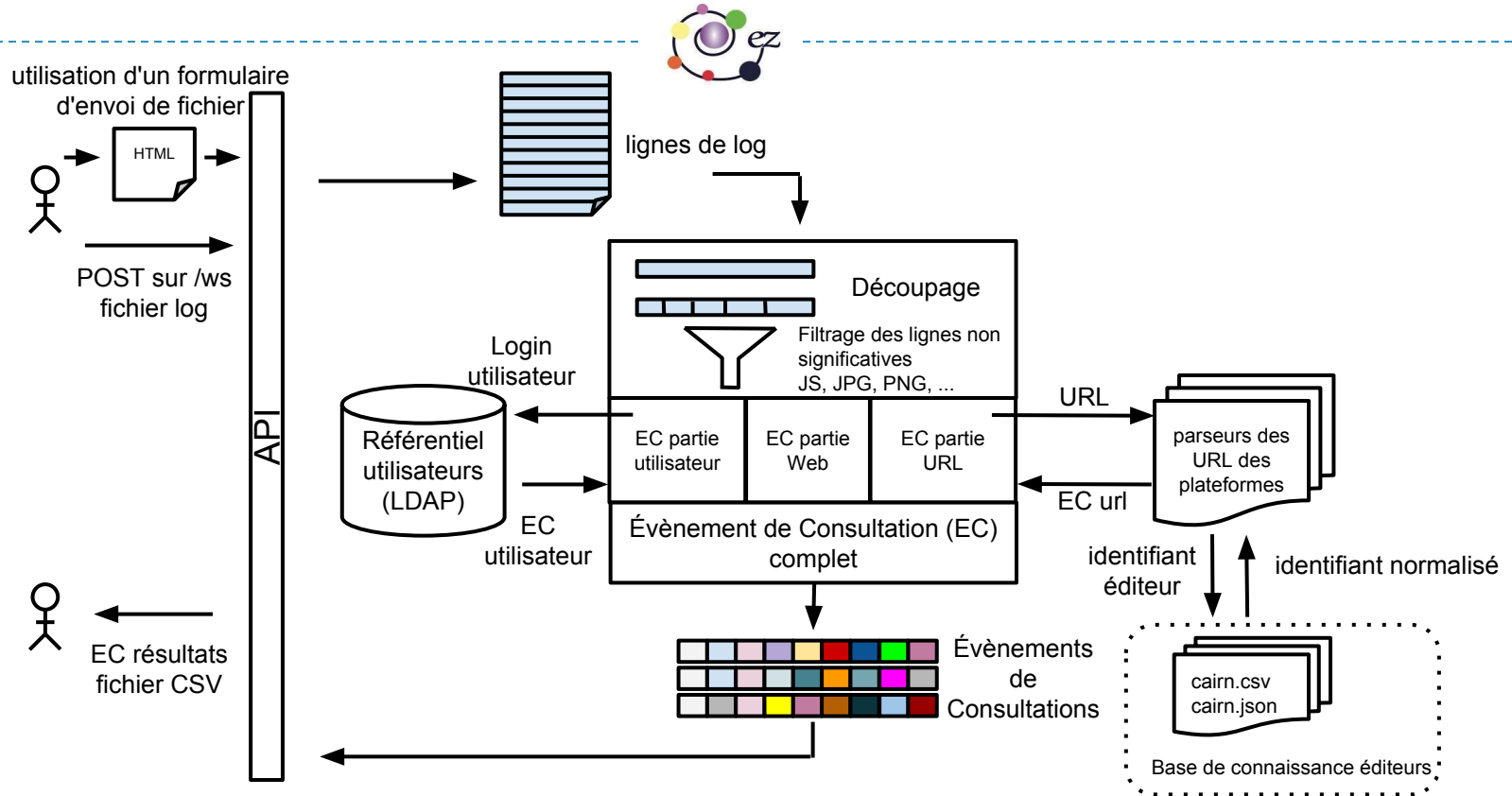
instance nationale +  
espace collaboratif



# Exemple d'utilisation d'ezPAARSE



# Architecture



# Coeur d'ezPAARSE : les parseurs



- Historique
- Démarche projet
- Présentation d'ezPAARSE/AnalogIST
- **Coeur d'ezPAARSE : les parseurs**
- Exploitation des résultats
- Écosystème d'ezPAARSE

# Logs générés par EZProxy



126.10.7.11 – stephane.gully@inist.fr [13/Mar/2012:16:10:01 +0100] "GET http://sciencedirect.com HTTP/1.1" 200 1993

126.10.7.11 – stephane.gully@inist.fr [13/Mar/2012:16:13:41 +0100] "GET http://sciencedirect.com/00014575/01.pdf HTTP/1.1" 200 465935



ISSN = 0001-4575  
id PDF = 01



# Structure d'une URL



[http://pdn.sciencedirect.com/science?  
\\_ob=MiamilimageURL&\\_cid=271664&\\_user=4046427&\\_pi  
i=S0001457512000747&\\_check=y&\\_origin=browse&\\_zon  
e=rslt\\_list\\_item&\\_coverDate=2012-07-31&wchp=dGLbVlt-  
zSkWb&md5=f5d8d157ccda6d597cb466af123dbff3/1-s2.  
0-S0001457512000747-main.pdf](http://pdn.sciencedirect.com/science?_ob=MiamilimageURL&_cid=271664&_user=4046427&_pii=S0001457512000747&_check=y&_origin=browse&_zone=rslt_list_item&_coverDate=2012-07-31&wchp=dGLbVlt-zSkWb&md5=f5d8d157ccda6d597cb466af123dbff3/1-s2.0-S0001457512000747-main.pdf)

# Structure d'une URL



http://pdn.sciencedirect.com/science?  
\_ob=MiamilImageURL&\_cid=271664&\_user=4046427&\_pi  
i=S**00014575**12000747&\_check=y&\_origin=browse&\_zon  
e=rslt\_list\_item&\_coverDate=2012-07-31&wchp=dGLbVlt-  
zSkWb&md5=f5d8d157ccda6d597cb466af123dbff3/1-s2.  
0-S**00014575**12000747-main.**pdf**



ISSN & type du fichier consulté

# Structure d'une URL



<http://www.sciencedirect.com/science/journal/00014575>



ISSN

En essayant l'URL manuellement on voit que c'est un sommaire HTML

# Structure d'une URL



[http://www.cairn.info/load\\_pdf.php?ID\\_ARTICLE=RFG\\_218\\_0009](http://www.cairn.info/load_pdf.php?ID_ARTICLE=RFG_218_0009)



Type PDF & ID éditeur  
Nécessite une table de  
correspondance pour  
décoder.

Code éditeur	ISSN
<b>RFG</b>	<b>0338-4551</b>
LMS	0027-2671

# Découper l'URL



http://pdn.sciencedirect.com/science?  
\_ob=MiamilImageURL&\_cid=271664&\_user=4046427&\_pii=**S00014575**12000747&\_check=y&\_origin=browse&\_zone=rslt\_list\_item&\_coverDate=2012-07-31&wchp=dGLbVlt-zSkWb&md5=f5d8d157ccda6d597cb466af123dbff3/1-s2.0-S0001457512000747-main.pdf

```
/_pii=S([0-9]{0,7}[0-9X])/i
```

# Une structure d'URL par plate-forme



- Sciencedirect = une structure d'URL
- Cairn = une structure d'URL
- ...
  
- Chaque plate-forme possède sa structure d'URL
- Écriture d'un parseur par plate-forme

# Limites de l'analyse des URLs



URL opaques (session, cryptages)  
Bases de connaissances à compléter manuellement

Exemple chez Springer :

<http://www.springerlink.com/content/j5q872410p510m63/fulltext.pdf>

Exemple chez Cairn :

[http://www.cairn.info/load\\_pdf.php?ID\\_ARTICLE=RFG\\_218\\_0009](http://www.cairn.info/load_pdf.php?ID_ARTICLE=RFG_218_0009)

# Code source d'un parseur



```
#!/usr/bin/env node
'use strict';
var Parser      = require('../.lib/parser.js');

module.exports = new Parser(function analyseUrl(parsedUrl) {
  var result = {};
  var param  = parsedUrl.query || {};
  var url    = parsedUrl.href;
  var match;

  switch (param._ob) {
    case 'MiamiImageURL':
      if (param._pii) {
        if ((match = /S{[0-9]{4}}{[0-9]{3}}{[0-9Xx]}{[0-9A-Za-z]*}/.exec(param._pii)) !== null) {
          result.unitid = 'S' + match[1] + match[2] + match[3];
          result.pissn  = match[1] + '-' + match[2];
          result.rtype  = 'ARTICLE';
          result.mime   = 'PDF';
        } else if ((match = /B{[0-9]{12}}{[0-9Xx]}/.exec(param._pii)) !== null) {
          result.unitid = 'B' + match[1] + match[2];
          result.pid    = match[1] + match[2];
          result.rtype  = 'BOOK';
          result.mime   = 'PDF';
        } else {
```



# Une interface ou des traitements planifiés



VIA un formulaire Web



- Contactez-nous par e-mail
- Visitez notre plateforme AnalogiST
- Consultez la documentation d'ezPAARSE
- Suivez-nous sur Facebook
- Suivez-nous sur Twitter
- Suivez-nous sur Google+

Traiter vos fichiers de log

Select fichiers: Aucun fichier choisi

Glissez ici

Options avancées

Traiter



- Contactez-nous par e-mail
- Visitez notre plateforme AnalogiST
- Consultez la documentation d'ezPAARSE
- Suivez-nous sur Facebook
- Suivez-nous sur Twitter
- Suivez-nous sur Google+

Le traitement de vos logs s'est terminé avec succès

100%

Télécharger le résultat en csv

Fichiers traités

fedes bibliovne eaproya 2013.05.20 log

État du traitement

Lignes en entrée	638721
ECs générés	8429
Plateformes reconnues	6
PDF	1854
HTML	3198
Taux de filtrage	97.88 %
Durée du traitement	15 m 15 s
Vitesse du traitement	695 lignes/s

En ligne de commande (cURL)

```
ubuntu@ubuntu: ~/ezpaarse
ubuntu@ubuntu:~/ezpaarse$ curl -X POST http://127.0.0.1:59599 -v --proxy "" --no-buffer
--data-binary @./test/dataset/sd.2012-11-30.300.log 2> /dev/null
domain;rtype;mime;doi;issn;elssn;platform;unitid;host;login;date:url;status;size
www.sciencedirect.com;ARTICLE;HTML;;1046-2023;;sd;;14.8.136.65;MAGDALEN;2012-11-30;http://www.sciencedirect.com:80/science/article/pii/S1046202310000411;200;114546
www.sciencedirect.com;ARTICLE;HTML;;1360-1385;;sd;;216.203.82.104;NICK;2012-11-30;http://www.sciencedirect.com:80/science/article/pii/S1360138512001719;200;206621
www.sciencedirect.com;ARTICLE;HTML;;1360-1385;;sd;;131.159.252.196;NICK;2012-11-30;http://www.sciencedirect.com:80/science/article/pii/S1360138512001513;200;171260
www.sciencedirect.com;ARTICLE;HTML;;1360-1385;;sd;;168.98.97.14;NICK;2012-11-30;http://www.sciencedirect.com:80/science/article/pii/S1360138512001689;200;103423
www.sciencedirect.com;ARTICLE;HTML;;0166-445X;;sd;;192.21.88.49;DARYL;2012-11-30;http://www.sciencedirect.com:80/science/article/pii/S0166445X08000027;200;151865
www.sciencedirect.com;ARTICLE;HTML;;0166-445X;;sd;;192.21.88.49;DARYL;2012-11-30;http://www.sciencedirect.com:80/science/article/pii/S0166445X03001772;200;171162
www.sciencedirect.com;ARTICLE;HTML;;0006-291X;;sd;;50.226.74.76;CECIL_MEDHURST;2012-11-30;http://www.sciencedirect.com:80/science/article/pii/S0006291X0202003X;200;126031
www.sciencedirect.com;ARTICLE;HTML;;0166-445X;;sd;;192.21.88.49;DARYL;2012-11-30;http://www.sciencedirect.com:80/science/article/pii/S0166445X11000154;200;227805
www.sciencedirect.com;TOC;MISC;;1087-0792;;sd;;96.178.240.98;MALIKA.HAHN;2012-11-30;http://www.sciencedirect.com:80/science/journal/10870792;200;80713
www.sciencedirect.com;ARTICLE;HTML;;0166-445X;;sd;;192.21.88.49;DARYL;2012-11-30;http://www.sciencedirect.com:80/science/article/pii/S0166445X03001772;200;13444
www.sciencedirect.com;ARTICLE;HTML;;0166-445X;;sd;;192.21.88.49;DARYL;2012-11-30;http://www.sciencedirect.com:80/science/article/pii/S0166445X03001772;200;171154
www.sciencedirect.com;ARTICLE;HTML;;0048-9697;;sd;;192.21.88.49;DARYL;2012-11-30;http://www.sciencedirect.com:80/science/article/pii/S0048969703000469;200;133861
www.sciencedirect.com;ARTICLE;HTML;;1532-0456;;sd;;192.21.88.49;DARYL;2012-11-30;http://www.sciencedirect.com:80/science/article/pii/S1532045605000979;200;165718
www.sciencedirect.com;TOC;MISC;;1087-0792;;sd;;96.178.240.98;MALIKA.HAHN;2012-11-30;http://www.sciencedirect.com:80/science/journal/10870792;200;80712
www.sciencedirect.com;ARTICLE;HTML;;0166-445X;;sd;;192.21.88.49;DARYL;2012-11-30;http://www.sciencedirect.com:80/science/article/pii/S0166445X02000498;200;174462
www.sciencedirect.com;TOC;MISC;;1389-9457;;sd;;96.178.240.98;MALIKA.HAHN;2012-11-30;http://www.scienc
```

# Exploitation des résultats



- Historique
- Démarche projet
- Présentation d'ezPAARSE/AnalogIST
- Coeur d'ezPAARSE : les parseurs
- **Exploitation des résultats**
- Écosystème d'ezPAARSE

# ezPAAARSE : exploitation des résultats



```
fede.bibliovie.ezproxy.2013.11.20.log x
com:80/gadgets/services/gadgets/v6/css/gadgets_ext/gadgets/toolbar/SDArticleGadgetOverRide, jquery
ui-1.8.4.custom.css HTTP/1.1" 200 27584
690 APointe-a-Pitre-551-1-48-61.w80-8.abo.wanadoo.fr - 13SDVUMR6553 [20/Nov/2013:00:16:11 +0100]
"POST http://www.sciencedirect.com:80/science/dc/article/all HTTP/1.1" 200 1632
691 APointe-a-Pitre-551-1-48-61.w80-8.abo.wanadoo.fr - 13SDVUMR6553 [20/Nov/2013:00:16:11 +0100]
"GET http://www.sciencedirect.com:80/science/frag/50016706100000860/30b35129f7454a9ac367217015b
fb552bb7606a297de3c1873239a77c297f7ed4b2086eeeca3d586205a2eb73c06004f348aeef14811d6412b0b178477
b5e0d62e6c4e43db40c1563b468e2ca066d7f49fcc80eca0718ac9f3de894251c948320e16e4fe2750a46b6e6e4ef6a
94fa159ba39e98c88f8c7246b64de3a3be7ff23fa38e315200d0c0141ab1313f3f7a88226fab51335456b0e9bfe29f
77b0845bac1d7b909778a4214cdfef5cfd76ac383326bbdbed6e9d2e/frag HTTP/1.1" 200 13168
692 APointe-a-Pitre-551-1-48-61.w80-8.abo.wanadoo.fr - 13SDVUMR6553 [20/Nov/2013:00:16:12 +0100]
"GET http://www.sciencedirect.com:80/science/outlineJson/S0016706100000860/271789 HTTP/1.1"
200 4351
693 ds1b-094-223-089-139.pools.arcor-ip.net - 13SDVUMR5023 [20/Nov/2013:00:16:12 +0100] "GET http
://gate1.inist.fr:50162/login?url=http://www.nature.com/ng/ HTTP/1.1" 302 0
694 APointe-a-Pitre-551-1-48-61.w80-8.abo.wanadoo.fr - 13SDVUMR6553 [20/Nov/2013:00:16:12 +0100]
"GET http://www.sciencedirect.com:80/science/frag/50016706100000860/30b35129f7454a9ac367217015b
fb552bb7606a297de3c1873239a77c297f7ed4b2086eeeca3d586205a2eb73c06004f348aeef14811d6412b0b178477
b5e0d62e6c4e43db40c1563b468e2ca066d7f49fcc80eca0718ac9f3de894251c948320e16e4fe2750a46b6e6e4ef6a
94fa159ba39e98c88f8c7246b64de3a3be7ff23fa38e315200d0c0141ab1313f3f7a88226fab51335456b0e9bfe29f
77b0845bac1d7b909778a4214cdfef5cfd76ac383326bbdbed6e9d2e/frag 1 HTTP/1.1" 200 16661
695 81-64-144-82.rev.numericable.fr - 13SDVUMR8113 [20/Nov/2013:00:16:12 +0100] "GET http://www.
sciencedirect.
com:80/science/article/pii/S1934590911004851/pdf?ftid=5973fb12adbb182f72d246b39a9e46e&pid=1-
s2.0-S1934590911004851-main.pdf HTTP/1.1" 302 532
696 81-64-144-82.rev.numericable.fr - 13SDVUMR8113 [20/Nov/2013:00:16:12 +0100] "GET http://www.
sciencedirect.com:80/science/suggestedArt/citelist/pii/S1934590911004851/eid/1-s2.0-
S1934590911004851/serial HTTP/1.1" 200 6292
697 APointe-a-Pitre-551-1-48-61.w80-8.abo.wanadoo.fr - 13SDVUMR6553 [20/Nov/2013:00:16:12 +0100]
"GET http://www.sciencedirect.com:80/science/outlineJson/S0016706100000860/271789 HTTP/1.1"
200 4350
698 ds1b-094-223-089-139.pools.arcor-ip.net - 13SDVUMR5023 [20/Nov/2013:00:16:13 +0100] "GET http
://www.nature.com:80/ng/index.html HTTP/1.1" 200 50528
699 APointe-a-Pitre-551-1-48-61.w80-8.abo.wanadoo.fr - 13SDVUMR6553 [20/Nov/2013:00:16:13 +0100]
"GET http://www.sciencedirect.
com:80/science/UserActions/article/rightPane/serial/displayRightPanePanel/citeRelatedArt HTTP/1
.1" 200 413
700 APointe-a-Pitre-551-1-48-61.w80-8.abo.wanadoo.fr - 13SDVUMR6553 [20/Nov/2013:00:16:13 +0100]
"POST http://www.sciencedirect.com:80/science/reLinkAjaxReq_eid=1-s2.0-S0016706100000860
HTTP/1.1" 200 413
701 ds1b-094-223-089-139.pools.arcor-ip.net - 13SDVUMR5023 [20/Nov/2013:00:16:13 +0100] "GET http
://www.nature.com:80/common/includes/footer/03/style/footer.css HTTP/1.1" 200 294
702 ds1b-094-223-089-139.pools.arcor-ip.net - 13SDVUMR5023 [20/Nov/2013:00:16:13 +0100] "GET http
://www.nature.com:80/common/includes/footer/03/style/footer.css HTTP/1.1" 200 8996
703 ds1b-094-223-089-139.pools.arcor-ip.net - 13SDVUMR5023 [20/Nov/2013:00:16:13 +0100] "GET http
://www.nature.com:80/common/style/global_message.css HTTP/1.1" 200 2939
704 ds1b-094-223-089-139.pools.arcor-ip.net - 13SDVUMR5023 [20/Nov/2013:00:16:13 +0100] "GET http
```

0246ccf0-538f-11e3-82cd-f1c6e79ec852.csv - LibreOffice Calc

Libération Sans 10

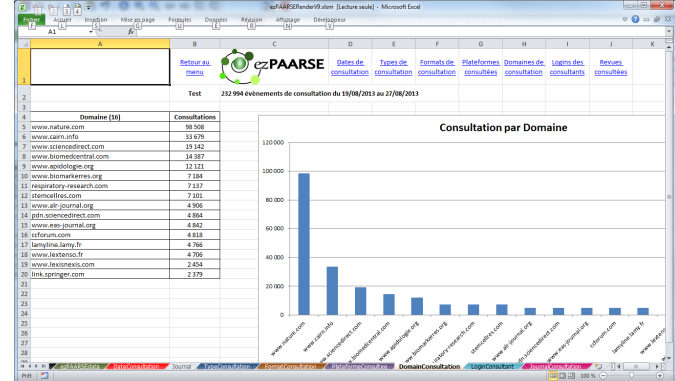
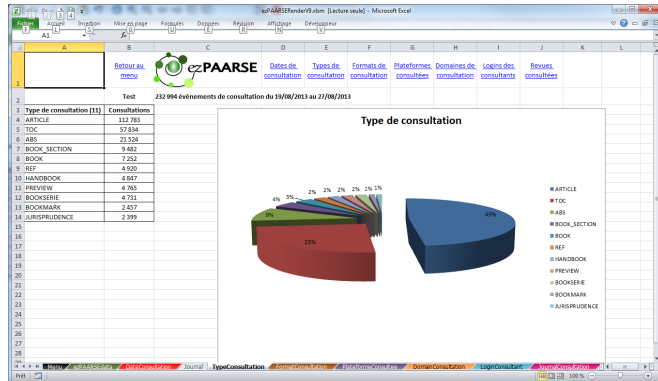
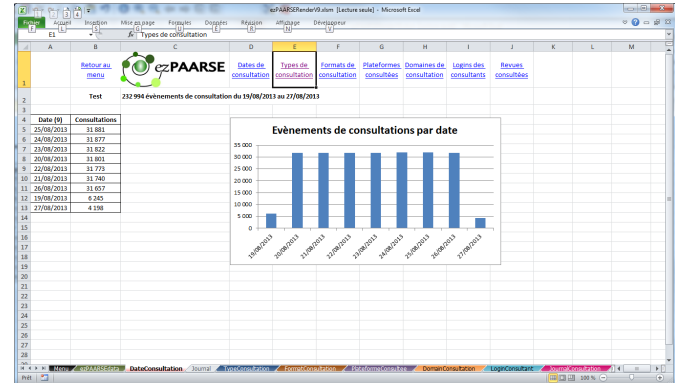
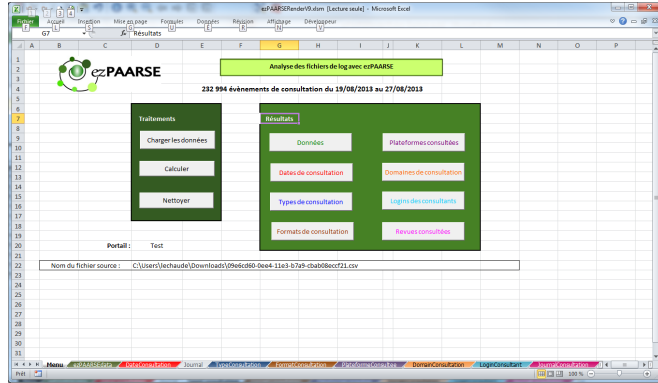
A1:AMJ1

	A	B	C	D	E	F	G	H
1	domain	type	mime	doi	piissn	eissn	title	platform
2	www.nature.com	TOC	MISC					ngp
3	www.nature.com	TOC	MISC		1474-1733	1474-1741	Nature Reviews Immunology	ngp
4	www.sciencedirect.com	ARTICLE	HTML		1931-3128			sd
5	www.sciencedirect.com	ARTICLE	HTML		0966-842X			sd
6	www.sciencedirect.com	ARTICLE	HTML		1471-4906			sd
7	www.nature.com	ARTICLE	HTML		1740-1526	1740-1534	Nature Reviews Microbiology	ngp
8	www.sciencedirect.com	ARTICLE	HTML		1534-5807			sd
9	onlineibrary.wiley.com	TOC	MISC			1365-3040		wiley
10	www.sciencedirect.com	TOC	MISC		0003-9861			sd
11	www.sciencedirect.com	ARTICLE	HTML		0304-4165			sd
12	www.sciencedirect.com	ARTICLE	HTML		0003-9861			sd
13	onlineibrary.wiley.com	TOC	MISC			1365-3040		wiley
14	onlineibrary.wiley.com	ARTICLE	PDF	10.1111/(ISSN)1365-3040	0140-7791	1365-3040	PLANT CELL & ENVIRONMENT	wiley
15	www.nature.com	ARTICLE	PDF		1740-1526	1740-1534	Nature Reviews Microbiology	ngp
16	www.sciencedirect.com	ARTICLE	HTML		0939-6411			sd
17	www.nature.com	ARTICLE	HTML		0018-067X	1365-2540	Heredity	ngp
18	link.springer.com	ABS	MISC	10.1007/s13238-013-3057-2	1674-800X	1674-8018	Protein & Cell	springer
19	www.sciencedirect.com	ARTICLE	HTML		1474-4422			sd
20	www.sciencedirect.com	ARTICLE	HTML		1474-4422			sd
21	www.sciencedirect.com	ARTICLE	HTML		0005-2728			sd
22	link.springer.com	TOC	MISC		0722-4028	1432-0975	Coral Reefs	springer
23	link.springer.com	ABS	MISC	10.1007/s00338-012-0991-2	0722-4028	1432-0975	Coral Reefs	springer
24	link.springer.com	ARTICLE	HTML	10.1007/s00338-012-0991-2	0722-4028	1432-0975	Coral Reefs	springer
25	onlineibrary.wiley.com	ABS	MISC			1600-0854		wiley
26	www.sciencedirect.com	ARTICLE	HTML		0302-2838			sd
27	link.springer.com	TOC	MISC		0722-4028	1432-0975	Coral Reefs	sd
28	link.springer.com	ABS	MISC	10.1007/s00338-012-0916-x	0722-4028	1432-0975	Coral Reefs	springer
29	onlineibrary.wiley.com	ARTICLE	PDF	10.1002/(ISSN)1099-1166	0885-6230	1099-1166	INTERNATIONAL JOURNAL OF GERIATRIC	wiley
30	download.springer.com	ARTICLE	PDF					springer
31	www.sciencedirect.com	ARTICLE	HTML		1473-0502			sd
32	www.sciencedirect.com	ARTICLE	HTML		0022-5347			sd

Rechercher

Feuille 1 / 1 Par défaut

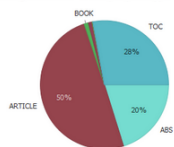
# Exploitation des résultats



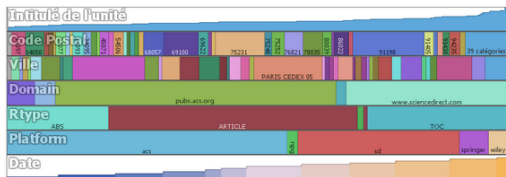
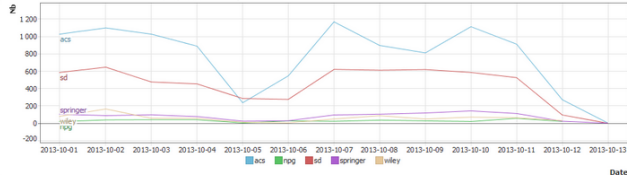
# Exploitation des résultats



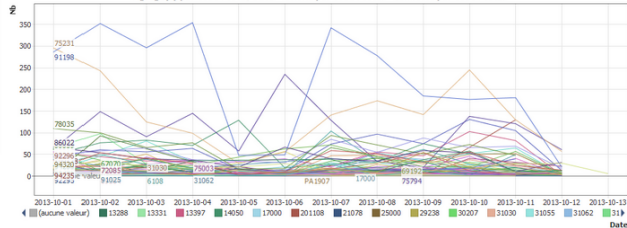
Nombre d'enregistrements Découper par RType Colorié par RType



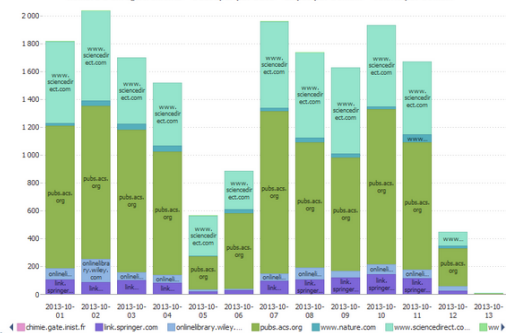
Agrégé(s) par Date, Platform Colorié par Platform Connécté par Platform



Agrégé(s) par Code Postal, Date Colorié par Code Postal Connécté par Code Postal



Nombre d'enregistrements Découper par Date Empilé par Domain Colorié par Domain



Nombre d'enregistrements

Platform, RType	2013-10-01	2013-10-02	2013-10-03	2013-10-04	2013-10-05	2013-10-06	2013-10-07	2013-10-08	2013-10-09	2013-10-10	2013-10-11	2013-10-12	2013-10-13	Tous
ABS	260	300	261	247	82	204	354	217	244	295	224	75	3	2 266
ARTICLE	424	477	430	337	81	194	468	387	300	400	351	107		4 034
BOOK_SECTION														5
TOC	342	321	336	305	74	149	347	293	257	348	335	90		3 197
Tous	1 026	1 099	1 027	889	237	547	1 169	897	811	1 113	912	272	3	10 002
ARTICLE	2	12	17	18		6		14	8	4	22	6		109
TOC	19	27	24	23	5	22	24	22	21	17	35	16		255
Tous	21	39	41	41	5	28	24	36	29	21	57	22		364
ARTICLE	456	514	341	339	230	214	461	432	507	434	389	85	1	4 401
BOOK														31
BOOKSERIE	3	1	3	2		1	2		4	8	3	2		34
HANDBOOK														3
Tous	474	534	344	341	231	215	463	436	511	442	392	87	1	4 344

# Écosystème d'ezPAARSE



- Historique
- Démarche projet
- Présentation d'ezPAARSE/AnalogIST
- Coeur d'ezPAARSE : les parseurs
- Exploitation des résultats
- **Écosystème d'ezPAARSE**

# Métiers et interactions



- Métiers
  - Informaticiens
  - Documentalistes
  - Négociateurs
- Interactions
  - Analyses des plateformes (D)
  - Développement des parseurs (I)
  - Production d'indicateurs (N)



# Besoins d'organisations



- Développement collaboratif multi-métiers
- Maintenance (prévisible donc doit être facilitée)
- Mutualisation (échelle internationale)
- Quelques réponses :
  - Méthode agile
  - Logiciel libre
  - Qualité logicielle (ex: architecture modulaire/parseur multi-langages)



# Logiciel libre



- Logiciel libre (licence CeCILL v2) : conçue conjointement par le CEA, le CNRS et l'INRIA, équivalent de la licence GPL en droit français
- Licence à caractère "copyleft" (liberté de modification et de redistribution) héréditaire
- Encourage et encadre ainsi les contributions dans un objectif de mutualisation (code et bases de connaissances).
- Partage / élargissement des partenariats
- Diffusion du code source à la communauté

# GitHub



- SGV, déploiement et intégration
- Visibilité (valorisation des dev)
- Workflow collaboratif
  - Compte organisation (droits d'accès)
  - Base de connaissance (pull request : parseur + pkb)
  - Authentification déléguée (couplage avec AnalogIST)
  - Issues (bugtracker)



**GitHub**

<https://github.com/ezparse-project/>

# Javascript (NodeJS)



ezPAARSE est écrit en javascript (NodeJS)

## NodeJS

- Javascript côté serveur
- Plate-forme polyvalente de développement d'applications réseau “scalable”
- Construit sur le V8 de Google
- Mono-thread
- Événementiel, I/O non-bloquant
- Gestion des “stream” facilitée

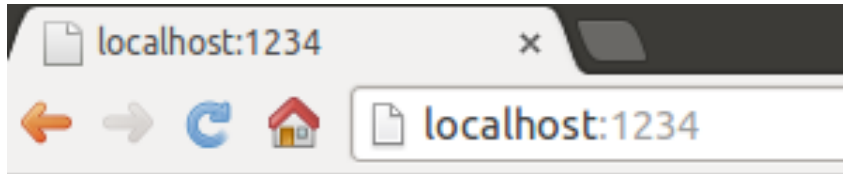


# NodeJS - Serveur web



```
app.js
1  var http = require('http');
2  http.createServer(function (req, res) {
3    res.writeHead(200, {'Content-Type': 'text/plain'});
4    res.end('Bonjour Montpellier !\n');
5  }).listen(1234, '127.0.0.1');
6  console.log('Serveur actif à http://127.0.0.1:1234/');
```

```
ubuntu@ubuntu:~/ezparse/tmp$ node app.js
Serveur actif à http://127.0.0.1:1234/
█
```



The screenshot shows a web browser window with the address bar set to 'localhost:1234'. The browser's navigation bar includes back, forward, refresh, and home buttons. The page content displays the text 'Bonjour Montpellier !' in a bold, black font.

Bonjour Montpellier !

# Bonnes pratiques



- Règles de codage (jshint)
- **TDD, + 70 tests (mocha)**
- **Documentation intégrée (beautiful docs)**
- Intégration continue (travis-ci)
- Scrum board (pivotal tracker)

# Tests unitaires avec Mocha

mocha

simple, flexible, fun



```
ubuntu@ubuntu:~/ezparse$ make test
```

```
.....  
75 passing (25s)
```

```
ubuntu@ubuntu:~/ezparse$
```

```
connection-test.js x  
1  /*jshint node: true, maxlen: 100, maxerr: 50, indent: 2 */  
2  /*global describe, it*/  
3  'use strict';  
4  
5  var helpers = require('./helpers.js');  
6  var appname = require('../package.json').name;  
7  
8  describe('The server', function () {  
9    it('sends a page containing the application name', function (done) {  
10     helpers.get('/', function (error, res, body) {  
11       if (error) {  
12         throw error;  
13       }  
14       if (!res) {  
15         throw new Error('The application is not running');  
16       }  
17       res.should.have.status(200);  
18       res.body.should.include(appname);  
19       done();  
20     });  
21   });  
22 });
```

```
ubuntu@ubuntu:~/ezparse$ make test-platforms-verbose
```

```
. The platform wiley is usable: 164ms  
. The platform springer is usable: 159ms  
. The platform sd is usable: 157ms  
. The platform persee is usable: 158ms  
. The platform npg is usable: 162ms  
. The platform muse is usable: 149ms  
. The platform metapress is usable: 155ms  
. The platform lextenso is usable: 149ms  
. The platform lexisnexis is usable: 150ms  
. The platform lamyline is usable: 161ms  
. The platform jstor is usable: 154ms  
. The platform gsw is usable: 154ms  
. The platform erudit is usable: 142ms  
. The platform edp is usable: 148ms  
. The platform dalloz is usable: 149ms  
. The platform cup is usable: 143ms  
. The platform cairn is usable: 150ms  
. The platform bmc is usable: 149ms  
. The platform ar is usable: 143ms  
. The platform acs is usable: 145ms
```

```
20 passing (3s)
```

```
ubuntu@ubuntu:~/ezparse$
```

# Documentation avec “Beautiful docs”



```
FOLDERS
  ezpaarse
    bin
    build
    doc
    images
    theme
    admin.md
    changelog.md
    developer.md
    doubleclick.md
    ec-attributes.md
    formats.md
    installation.md
    knowledge-base.md
    makefile.md
    manifest.json
    multi-os.md
    outputfields.md
    presentation.md
    qualification.md
    quickstart.md
    routes.md
    tools.md
    tree.md
    userfields.md

presentation.md * quickstart.md * developer.md *
1 # Introduction #
2
3 [ezPAARSE]() permet l'analyse, l'enrichissement et l'exploitation des
4 logs d'accès collectés localement aux ressources électroniques proposées
5 par les plate-formes Web des éditeurs.
6
7 Vous avez accès ici à la documentation locale du progiciel, pour une
8 documentation plus détaillée du projet se reporter à [la plateforme
9 AnalogIST](http://analogist.couperin.org)
```



- Contactez-nous par e-mail
- Visitez notre plateforme AnalogIST
- Consultez la documentation d'ezPAARSE
- Suivez-nous sur Facebook
- Suivez-nous sur Twitter
- Suivez-nous sur Google+



Search

## Introduction

ezPAARSE permet l'analyse, l'enrichissement et l'exploitation des logs d'accès collectés localement aux ressources électroniques proposées par les plate-formes Web des éditeurs.

Vous avez accès ici à la documentation locale du progiciel, pour une documentation plus détaillée du projet se reporter à [la plateforme AnalogIST](http://analogist.couperin.org)

### Démarrage rapide

- Pré-requis
- Installation
- Tester l'installation
- Usage
- Paramétrage avancé

### Installation

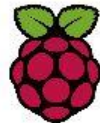
- Pré-requis
- Procédure d'installation

# Multi-plateforme



**ezPAARSE**

**node** JS™





# En conclusion, pourquoi les JRES ?



- **ezPAARSE c'est un logiciel libre**
  - Facile à installer et à tester (multi-plateforme)
  - Utilisant des technologies innovantes
- **Enjeux autour des ressources numériques**
  - Acquisitions par les bibliothèques
  - Connaissance des usages au niveau ESR
- **Nous avons besoin de vous ! (le réseau)**
  - Les documentalistes ont besoin des logs (admin sys)
  - Les contributions sont facilitées et attendues

# ezPAARSE



Des Questions ?

<http://ezpaarse.couperin.org>

<http://analogist.couperin.org>

<https://twitter.com/ezpaarse>

