

LAPAUZE Johann
NIVEAU Soline



Agrégation de flux RSS

Études d'Approfondissement Réseaux

6 Novembre 2009

Sommaire

- ▶ Standard RSS et Atom
 - ▶ Objectifs
 - ▶ Domaines d'utilisation
 - ▶ Technologies et protocoles
 - ▶ Limites
- ▶ Agrégateur RSS
 - ▶ Principes
 - ▶ Agrégateurs existants
- ▶ Conclusion
- ▶ Perspectives
- ▶ Démonstrations

Lexique

- ▶ RSS = Rich Site Summary
= RDF Site Summary
= Really Simple Syndication
- ▶ Flux RSS = Fil RSS = Canal RSS
- ▶ Agrégateur RSS = Lecteur RSS



1^{ère} Partie – Les Flux RSS

Pourquoi RSS ?

- ▶ Objectifs :
 - ▶ Avertir d'une mise à jour
 - ▶ Avertir d'une nouveauté
 - ▶ Avertir d'un nouvel article
- ▶ Outil déjà existant : La newsletter (SPAM)
- ▶ Avantages :
 - ▶ Présentation minimale des informations
 - ▶ Segmenter l'information en plusieurs thèmes

Exemples d'utilisation

- ▶ Flux sur l'actualité d'un site de e-Commerce :
 - ▶ Un flux sur les nouveaux produits
 - ▶ Un flux sur les promotions
 - ▶ Un flux sur les évènements
- ▶ Blog (Contenu sous forme de billets)
 - ▶ Nouveau billet disponible
 - ▶ Modification de billet
- ▶ Podcast
- ▶ Widget

Des origines aux évolutions

- ▶ Version 0.90 par Netscape (2000) : Trop complexe
- ▶ Version 0.91 par Userland Software : Trop limité
- ▶ Version 1.0 par RSS-DEV Working Group : Pour applications spécifiques
- ▶ Version 2.0 par Creative Commons (2002)

Qu'est ce qu'un flux RSS ?

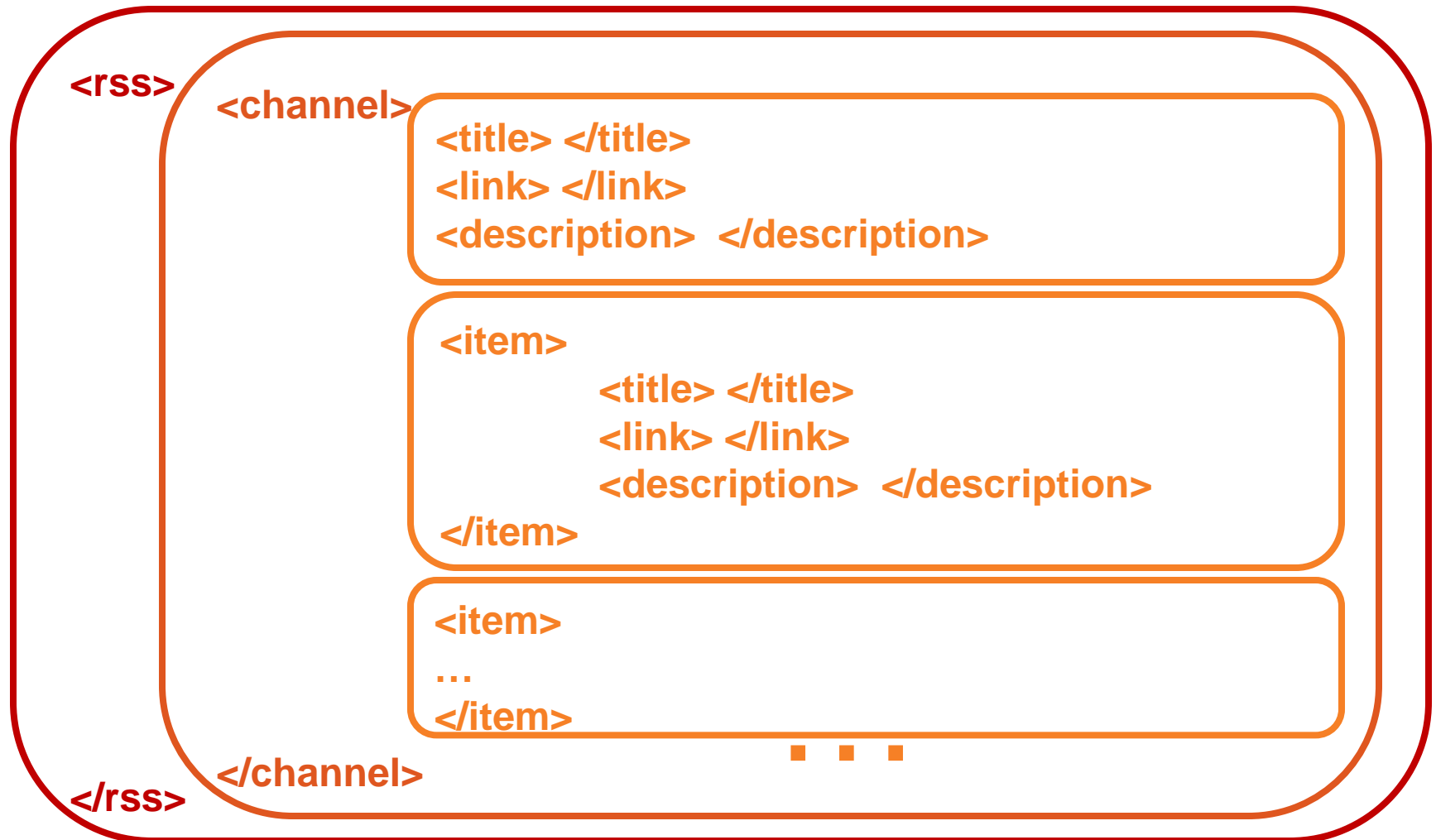
- ▶ *RSS : Really Simple Syndication*

RSS

XML

- ▶ Fichier XML respectant un formalisme et contenant :
 - ▶ Titre de l'information
 - ▶ Lien vers une page web
 - ▶ Brève description

Formalisme du fichier xml



Formalisme du fichier xml

Éléments obligatoires d'un <channel> :

- ▶ **<title>**: Nom du flux
 - ▶ **Convention** : Titre de la page web correspondante
- ▶ **<link>**: URL de la page web
- ▶ **<description>** : Brève description

Formalisme du fichier xml

Éléments optionnels d'un <channel> :

- ▶ <language> : Langage du flux
- ▶ <pubDate> : Date de publication
- ▶ <lastBuildDate> : Date de dernière modification
- ▶ <category>, <generator>, <image>...

Comment créer un flux RSS sur une page web ?

- ▶ Créer un fichier XML au format RSS
- ▶ Placer ce fichier dans l'architecture du site web
- ▶ Mettre la balise suivante pour que le browser détecte le flux
`<link rel="alternate" type="application/rss+xml" title="Flux RSS" href="flux.xml" />`

Flux RSS - Exemple

```
<rss version="2.0">
  <channel>
    <title>Mon site web</title>
    <link>http://monsiteweb.com</link>
    <description>Les news de mon site web</description>
    <item>
      <title>EaR Démo</title>
      <link>http://www.ear.com</link>
      <description>Présentation ear</description>
    </item>
  </channel>
</rss>
```

Limites du RSS 2.0

- ▶ Incompatibilités de tag
 - ▶ Pas de standard sur certaines balises

- ▶ Problème d'interopérabilité
 - ▶ Certaines informations ne seront pas prises en compte



2^{nde} Partie – Flux Atom

ATOM : une évolution du RSS

- ▶ Précise le type de contenu sans ambiguïté
- ▶ Types supportés variés (HTML, XHTML, XML)
- ▶ Références des documents variés (vidéo, audio..)

Atom - les versions

- ▶ Atom 0.2 (juillet 2003) – esquisse
 - ▶ Atom 0.3 (décembre 2003) - adopté par Google News
 - ▶ Atom 1.0 (2007) - normalisation
-

Atom - Introduction

- ▶ Un format destiné à la syndication de contenu
 - ▶ Un protocole de gestion de ressources
 - ▶ **Domaines d'application** : toute activité impliquant des mises à jour périodiques
-

Atom 1.0 : le format de diffusion

Le format du flux : XML

différences avec rss

	Atom	RSS 2.0	RSS 1.0
Conteneur	feed	channel	channel
Titre	title	title	title
URL	link	link	link
Sommaire	subtitle	description	description
Date	updated	lastBuildDate	dc:date
Logo	icon	image	image
Auteur	author	managingEditor	

Atom 1.0 : le format de diffusion

Le format des entrées : différences avec rss

	Atom	RSS 2.0	RSS 1.0
Conteneur	entry	item	item
Titre	title	title	title
URL	link	link	link
Description	summary	description	description
Date	updated	pubDate	dc:date
Image	logo		image

IETF (Internet Engineering Task Force) RFC 4287

AtomPub : le protocole

- ▶ Atom Publishing protocol ou AtomPub
 - ▶ Protocole HTTP
 - ▶ Un protocole de gestion de ressources
 - ▶ Création, suppression, mise à jour et publication
 - ▶ IETF (Internet Engineering Task Force) RFC 5023
-

AtomPub - gestion des ressources

- ▶ **Ressource** : objet accessible via internet, identifié par une @ IRI
- ▶ **Collection = fil Atom** : contient un ensemble de ressources :
 - ▶ Ressource d'entrée = entry d'Atom : identifiée par une adresse IRI = adresse internet
 - ▶ Ressource de média : ressource non représentée par un entrée Atom

Elles servent à organiser des ressources

AtomPub - document de service

```
<service>
  <workspace>
    <atom:title>Premier espace de travail</atom:title>
    <collection href="http://example.org/blog/main">
      <atom:title>Titre de la collection</atom:title>
      <categories href="http://example.com/cats/forMain.cats"/>
    </collection>
    <collection href="http://example.org/blog/pic">
      <atom:title>Pictures</atom:title>
      <accept>image/png</accept>
      <accept>image/jpeg</accept>
      <accept>image/gif</accept>
    </collection>
  </workspace>
</service>
```

AtomPub - scénario type client-serveur

- ▶ Client envoie requête GET à @Service
- ▶ Réponse du serveur (200 OK) : liste des collections disponibles (+ workspace)
- ▶ Client envoie requête GET sur @Collection
- ▶ Réponse du serveur (200 OK) : flux Atom
- ▶ Client crée nouvelle ressource : requête POST
- ▶ Réponse du serveur (201 Created) : entrée crée + @ressource crée
- ▶ Client envoie requêtes GET, PUT, DELETE pour gérer ressource

AtomPub : le protocole

▶ Utilisation de HTTP :

- ▶ GET : récupération d'une ressource
 - ▶ POST : création d'une nouvelle ressource
 - ▶ PUT : édition d'une ressource
 - ▶ DELETE : suppression d'une ressource
 - ▶ Serveur libre d'accepter/refuser/modifier requête
-



3^{ème} Partie – Agrégateurs RSS

Qu'est ce qu'un agrégateur RSS

- ▶ Permet de lire des flux RSS
- ▶ Centralise l'actualité de plusieurs sites web sans avoir à les consulter
- ▶ Types d'agrégateurs :
 - ▶ Applicatif
 - ▶ Online

Agrégateurs RSS ; S'abonner aux flux à travers :

- ▶ Une boîte mail comme Outlook
 - ▶ Alertes sous forme de mail
- ▶ Une application pour gérer les flux
 - ▶ HandyRSS pour Nokia
- ▶ Une page d'accueil de browser comme Netvibes
 - ▶ Alertes sous forme de widget

Agrégateurs RSS ; Récupérer les flux

► Utilisation du protocole HTTP

► RSS

Time	Source	Destination	Protocol	Info
1 0.000000	192.168.1.2	192.168.1.1	HTTP	GET /EaR/demo1/flux.xml HTTP/1.1
2 0.001714	192.168.1.1	192.168.1.2	HTTP/XML	HTTP/1.1 200 OK
3 0.002096	192.168.1.2	192.168.1.1	TCP	50513 > http-alt [ACK] seq=186 Ack=1257 win=65535 Len=0

► Atom

4 0.002398	192.168.1.2	192.168.1.1	HTTP	GET /EaR/demo2/atom.xml HTTP/1.1
5 0.004283	192.168.1.1	192.168.1.2	HTTP/XML	HTTP/1.1 200 OK
6 0.004621	192.168.1.2	192.168.1.1	TCP	50534 > http-alt [ACK] seq=186 Ack=1052 win=524280 Len=0

Podcast

- ▶ Flux RSS
- ▶ Plus seulement texte mais aussi contenu riche (vidéo, audio, présentations, animations)
- ▶ Téléchargement automatique des émissions audio ou vidéo sur ordinateurs/baladeurs
- ▶ iTunes, Juice, Netvibes, FeedMon (Windows)

Podcast

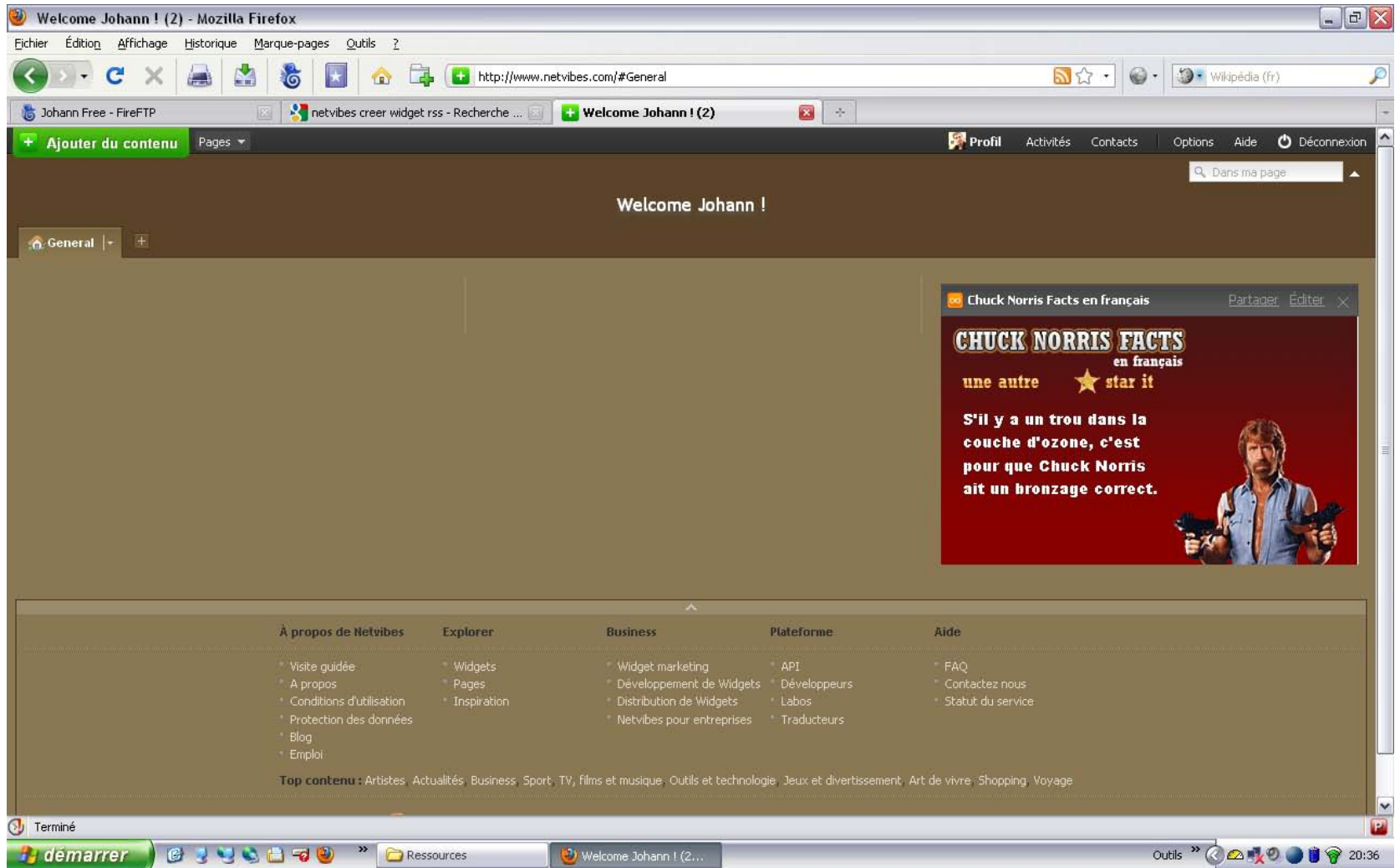
- ▶ HTTP pour les flux, et TCP pour le téléchargement

	Time	Source	Destination	Protocol	Info
4	0.000888	192.168.1.2	192.168.1.1	HTTP	GET /EaR/demo5/podcast.xml HTTP/1.1
6	0.007282	192.168.1.1	192.168.1.2	HTTP/XML	HTTP/1.1 200 OK
15	0.091457	192.168.1.2	192.168.1.1	HTTP	GET /EaR/demo5/Haven't_met_you_yet.mp3 HTTP/1.1
1632	5.662097	192.168.1.2	192.168.1.1	HTTP	GET /EaR/demo5/worrisome_heart.m4v HTTP/1.1

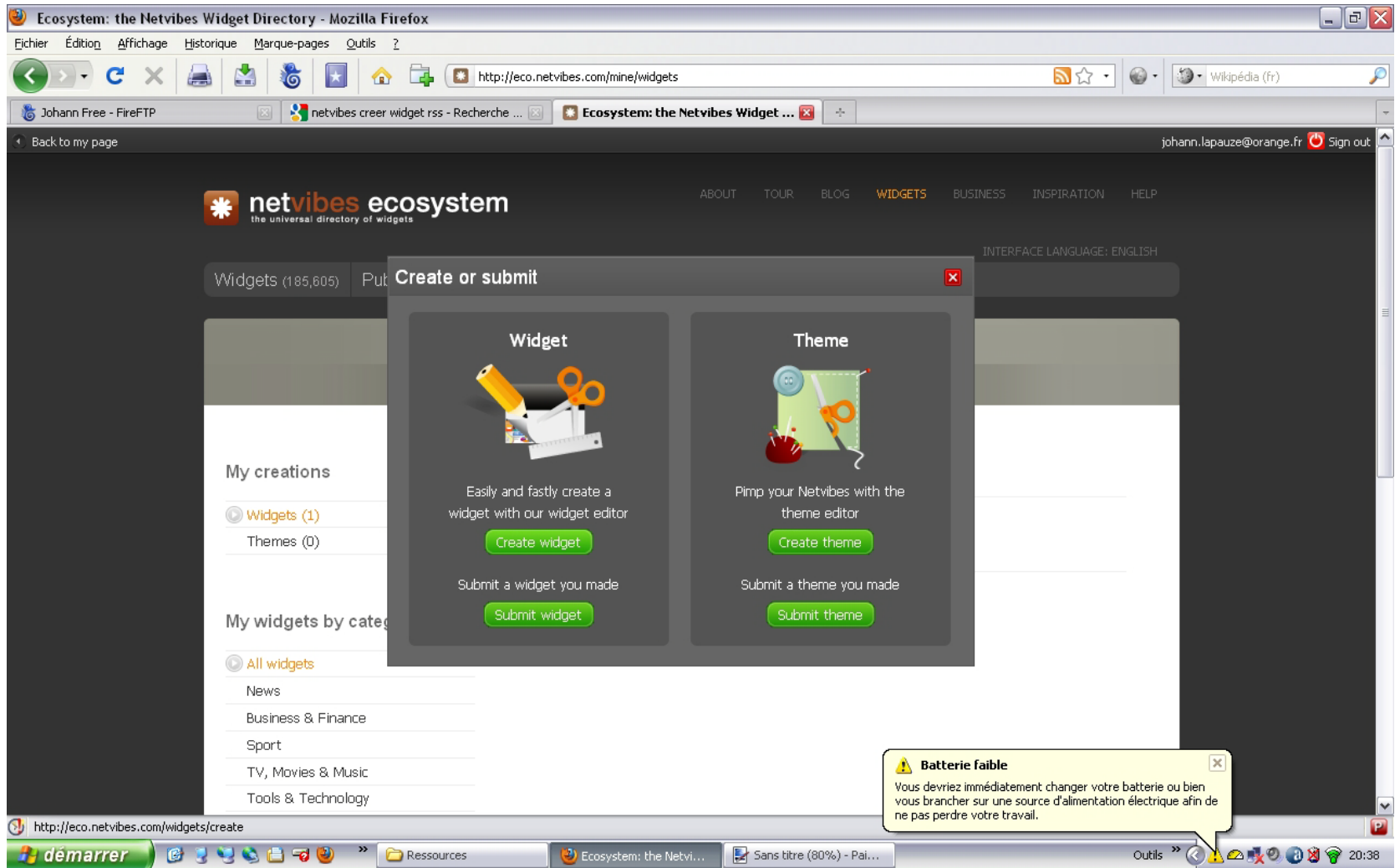
Widget pour iGoogle ou Netvibes

- ▶ Possibilité de créer des widgets via leurs sites
- ▶ Configuration du widget pour afficher les news du flux RSS
- ▶ Au chargement du widget, il met à jour son affichage avec les entrées du flux RSS

01 - Arrivée sur Netvibes



02 - Création d'un widget



03 - Choix d'un widget pour un Flux RSS

Ecosystem: the Netvibes Widget Directory - Mozilla Firefox

Fichier Édition Affichage Historique Marque-pages Outils ?

http://eco.netvibes.com/widgets/create

Johann Free - FireFTP netvibes creer widget rss - Recherche ... Ecosystem: the Netvibes Widget ...

Select the type of widget to create

Blog/News Widget

from an RSS, podcast feed

News - Sports

- Spurs Rule Third Quarter To Force Winner-Take-All - ESPN - 3 hours ago
- Injured shoulder sidelines Cavaliers' Daniel Gibson - Boston Herald - 1 hour ago
- Big Brown so good, winning likely a foregone conclusion - Fort Worth Star Telegram - 48 minutes ago
- Daboll doesn't recall talk - Boston Globe - 1 hour ago
- Flyers avoid elimination, beat Penguins, 4-2, in Game 4 - Philadelphia Daily News - 3 hours ago
- Giants' bullpen falters in 8-7 loss to Astros - San Jose Mercury News - 1 hour ago
- Mets' Wagner calls out teammates for ducking media - MLB - ESPN - 1 hour ago

Locate the feed address on the website.
It is usually represented with one of the following icons:

[Get more information about feeds.](#)

[Create a blog news widget](#)

Multifeed Widget

from a selection of RSS or podcast feeds

News - Sports

World U.S. Business Sports

Spurs Rule Third Quarter To Force Winner-Take-All - ESPN - 3 hours ago

Injured shoulder sidelines Cavaliers' Daniel Gibson - Boston Herald - 1 hour ago

next →

[Create a multifeed widget](#)

Batterie faible
Vous devriez immédiatement changer votre batterie ou bien vous brancher sur une source d'alimentation électrique afin de ne pas perdre votre travail.

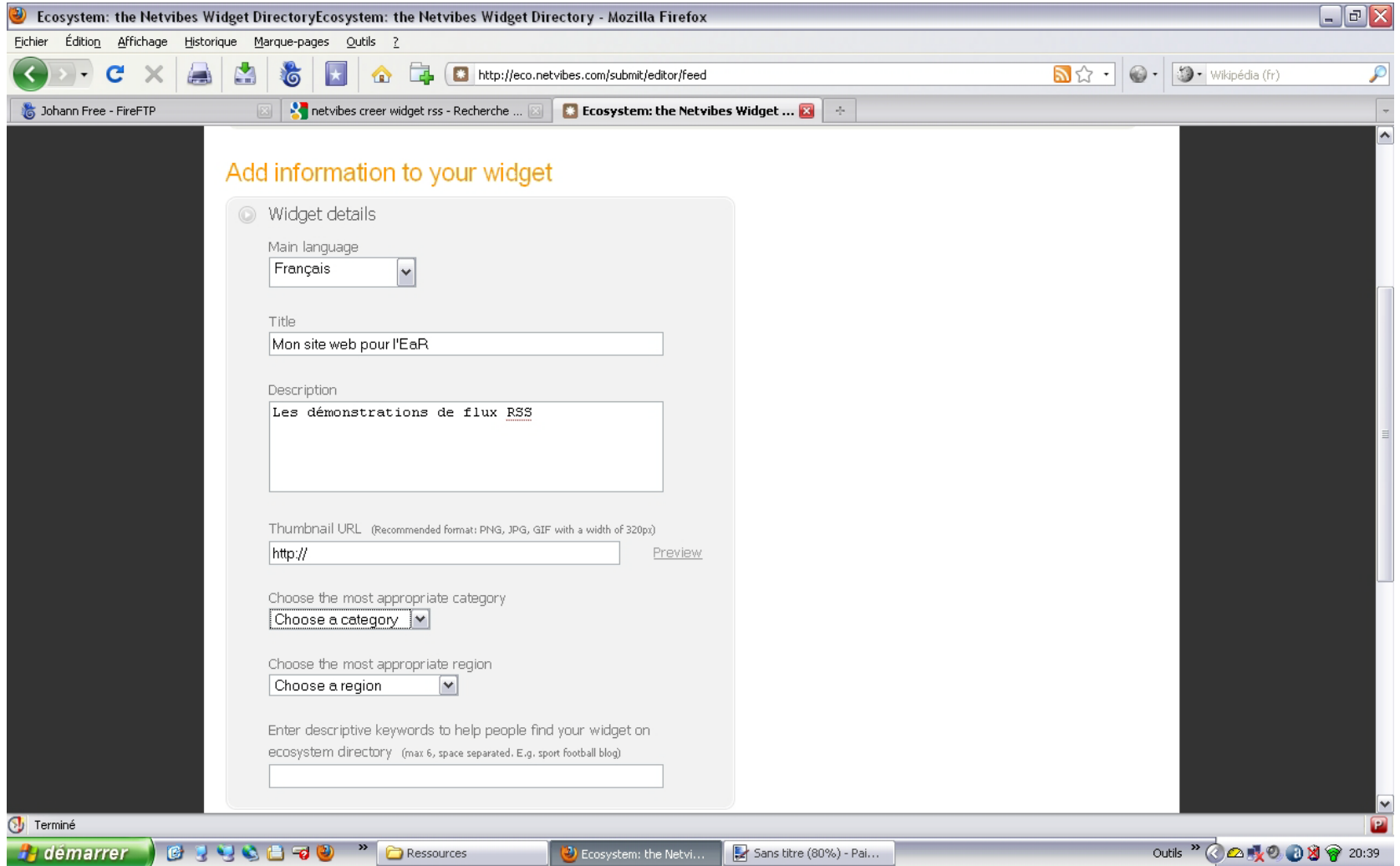
http://eco.netvibes.com/submit/editor/feed

démarrer Ressources Ecosystem: the Netvibes Sans titre (80%) - Pai... Outils 20:38

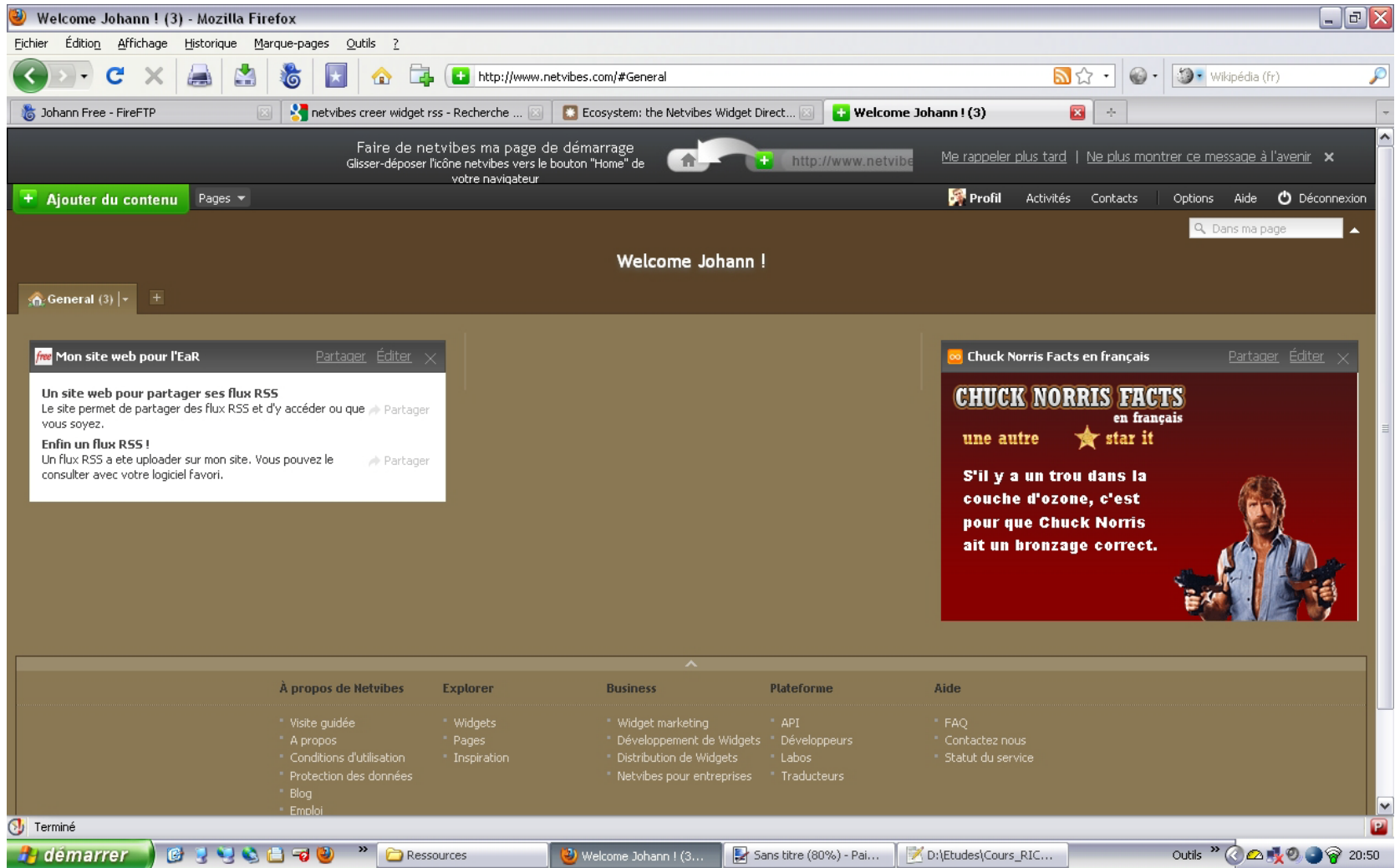
04 - Récupération du flux

The screenshot shows a Mozilla Firefox browser window displaying the 'Ecosystem: the Netvibes Widget Directory' website. The address bar shows the URL `http://eco.netvibes.com/submit/editor/feed`. The page features a navigation menu with links for ABOUT, TOUR, BLOG, WIDGETS, BUSINESS, INSPIRATION, and HELP. Below the navigation, there are statistics for Widgets (185,605), Public pages (31,087), Themes (1,114), and My creations (0), along with a 'Create or submit' button. A progress bar indicates four steps: 1. Select content, 2. Customize, 3. Add info, and 4. Submit & share. The current step is 'Submit a new feed', which includes instructions: 'Please provide a link to the feed you want widgetized. It can be either an RSS or an Atom feed. If you can't find the feed address, enter the website address instead. Netvibes will try to autodiscover it.' A text input field contains the URL `http://johann.lapauze.free.fr/EaR/flux.xml`. A 'Next step' button is visible at the bottom of the form. The browser's taskbar at the bottom shows the system tray with the time 20:38 and various icons.

05 - Paramétrage du widget



06 - Publication du widget



07 - Modification du flux et mise à jour du widget

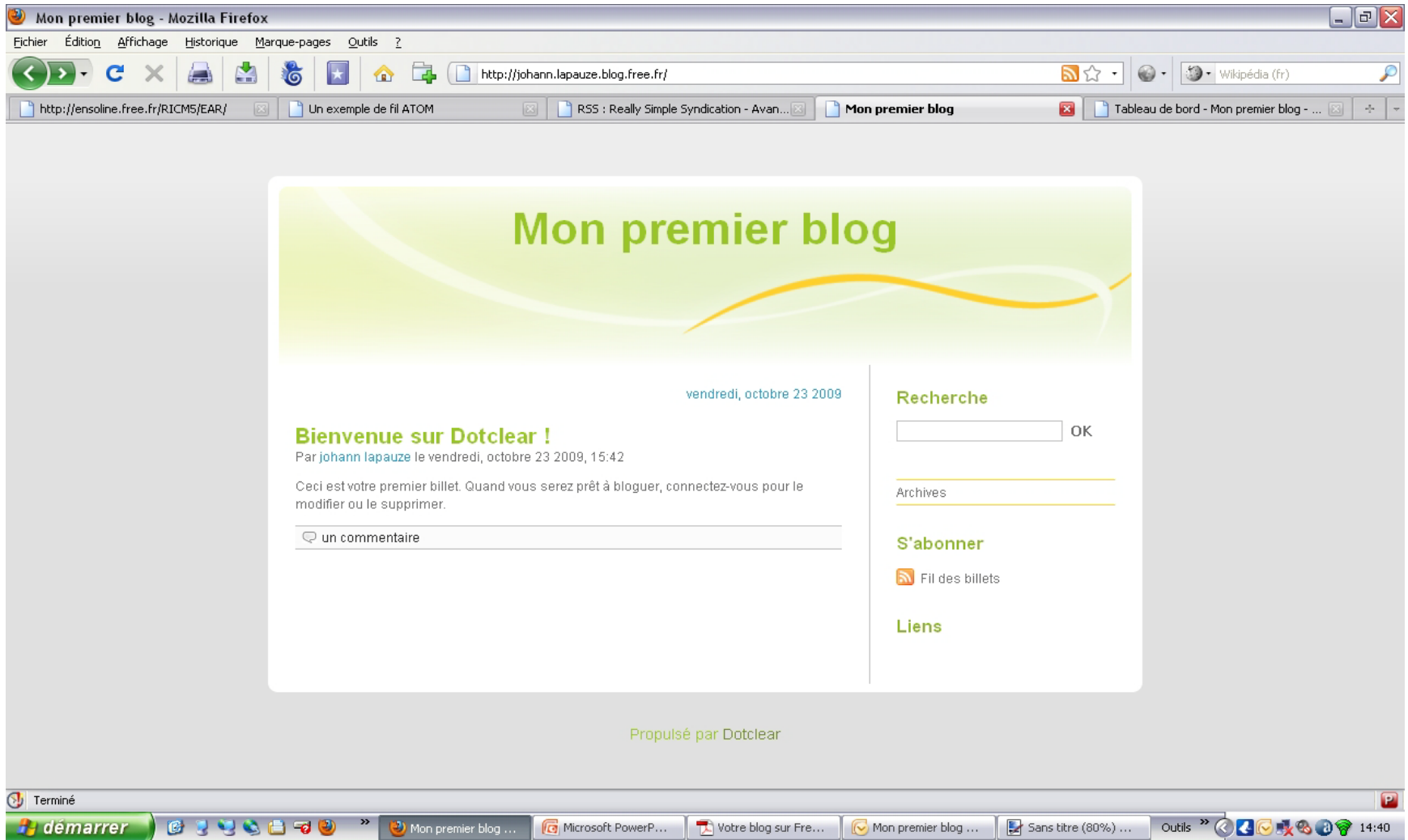
The screenshot shows a Mozilla Firefox browser window with the following elements:

- Browser Title Bar:** Welcome Johann ! (3) - Mozilla Firefox
- Address Bar:** http://www.netvibes.com/#General
- Navigation Bar:** Includes a home button and a message: "Faire de netvibes ma page de démarrage. Glisser-déposer l'icône netvibes vers le bouton 'Home' de votre navigateur."
- Dashboard Header:** "Welcome Johann !" with a search bar "Dans ma page".
- Widgets:**
 - Mon site web pour l'EaR:** A widget for sharing RSS feeds. It contains the text: "Un site web pour partager ses flux RSS. Le site permet de partager des flux RSS et d'y accéder ou que vous soyez." and "BLABLA blablablablablabalabala."
 - Chuck Norris Facts en français:** A widget with a red background and a photo of Chuck Norris. It says: "CHUCK NORRIS FACTS en français. une autre star it. S'il y a un trou dans la couche d'ozone, c'est pour que Chuck Norris ait un bronzage correct."
- Footer:** A navigation menu with categories: "À propos de Netvibes", "Explorer", "Business", "Plateforme", and "Aide".
- Windows Taskbar:** Shows the "démarrer" button, several icons, and open applications including "Ressources", "Welcome Johann ! (3...)", "Sans titre (80%) - Pai...", and "D:\Etudes\Cours_RIC...".

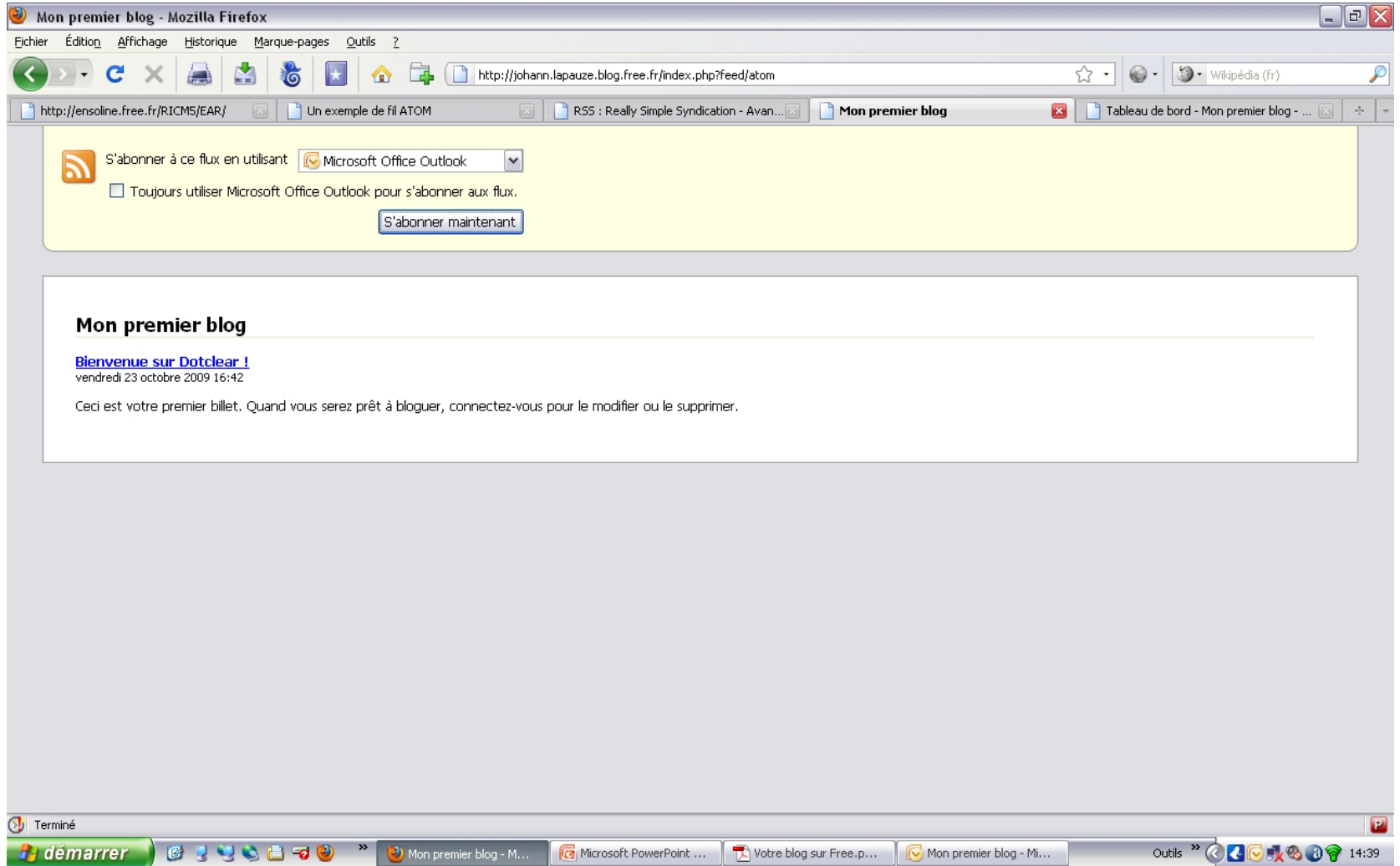
Blog

- ▶ Abonnement au flux Atom du blog
- ▶ Dès qu'un nouveau billet est publié, le flux est mis à jour
- ▶ Les clients abonnés sont avertis de la mise à jour

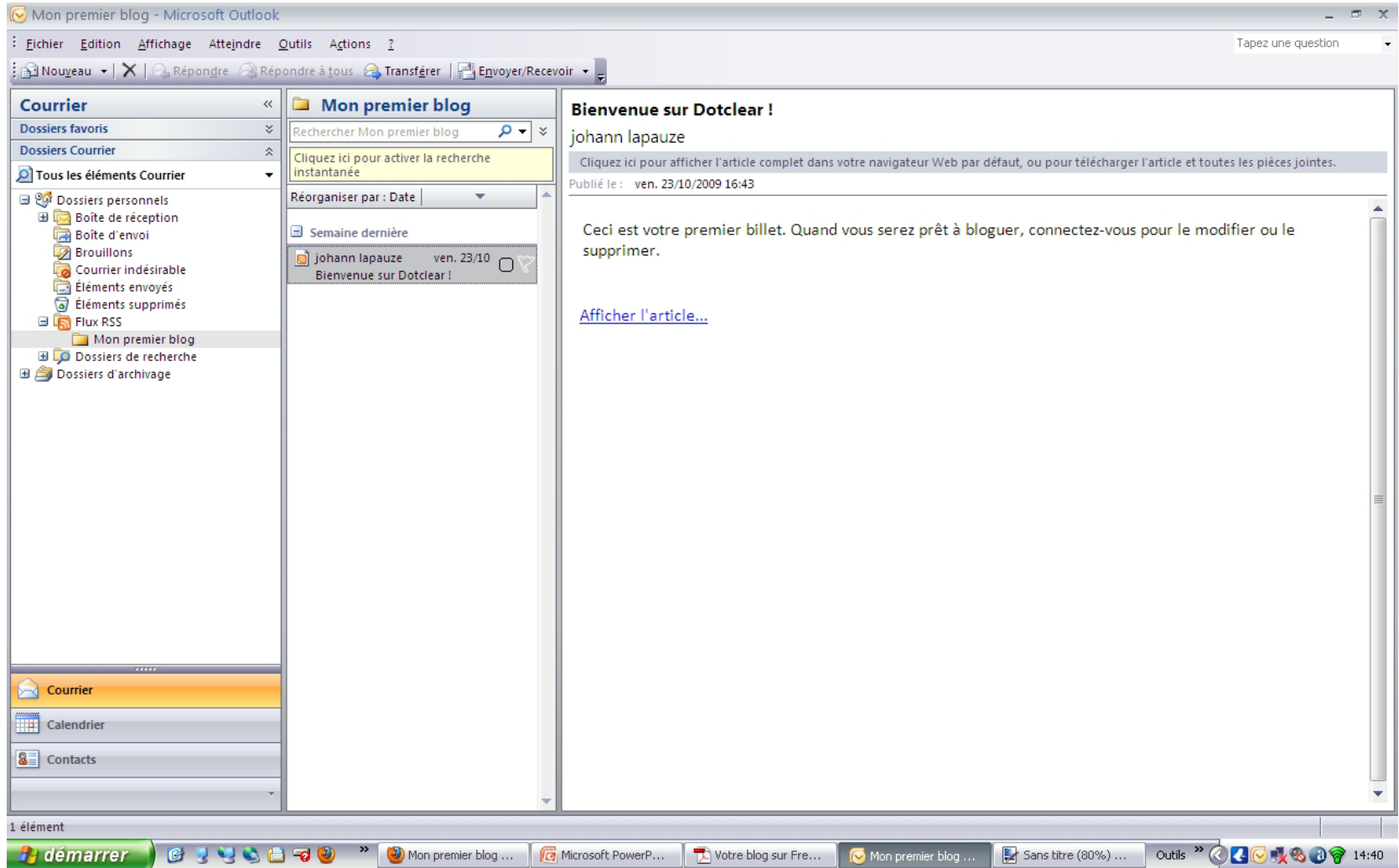
01 - Arrivée sur le Blog



02 - Demande d'abonnement



03 - Abonnement via Outlook



04 - Création d'un nouvel article

Tableau de bord - Mon premier blog - Dotclear - 2.1.5 - Mozilla Firefox

Fichier Édition Affichage Historique Marque-pages Outils ?

http://johann.lapauze.blog.free.fr/admin/index.php

http://ensoline.free.fr/RICMS/EAR/ Un exemple de fil ATOM RSS : Really Simple Syndication - Avan... Mon premier blog Tableau de bord - Mon premier ...

Catégories
Gestionnaire de médias
Pages
Tags
Widgets de présentation

EXTENSIONS
Antispam
Import/Export
Liens

Préférences utilisateur
Paramètres du blog
Apparence du blog
Pages
Liens

Billet rapide

* Titre :
Essai

* Contenu :
Mise à jour de mon blog !

Catégorie : [v]

enregistrer enregistrer et publier

Actualités

[Instal Party 4 \(le retour de la revanche\)](#)
27 octobre 2009: *Puisque je me suis fait balancer et qu'il est désormais officiel que j'ai la charge de l'événementiel Dotclear, je...*

[Du nouveau dans les versions de développement](#)
22 octobre 2009: *Trois semaines après la sortie de la dernière version, le boulot continue sur l'application : Si la...*

[Il n'y a pas que du code dans Dotclear](#)
15 octobre 2009

[Dotclear 2.1.6](#)
01 octobre 2009

[Dotclear 2.1.6 RC](#)
23 septembre 2009

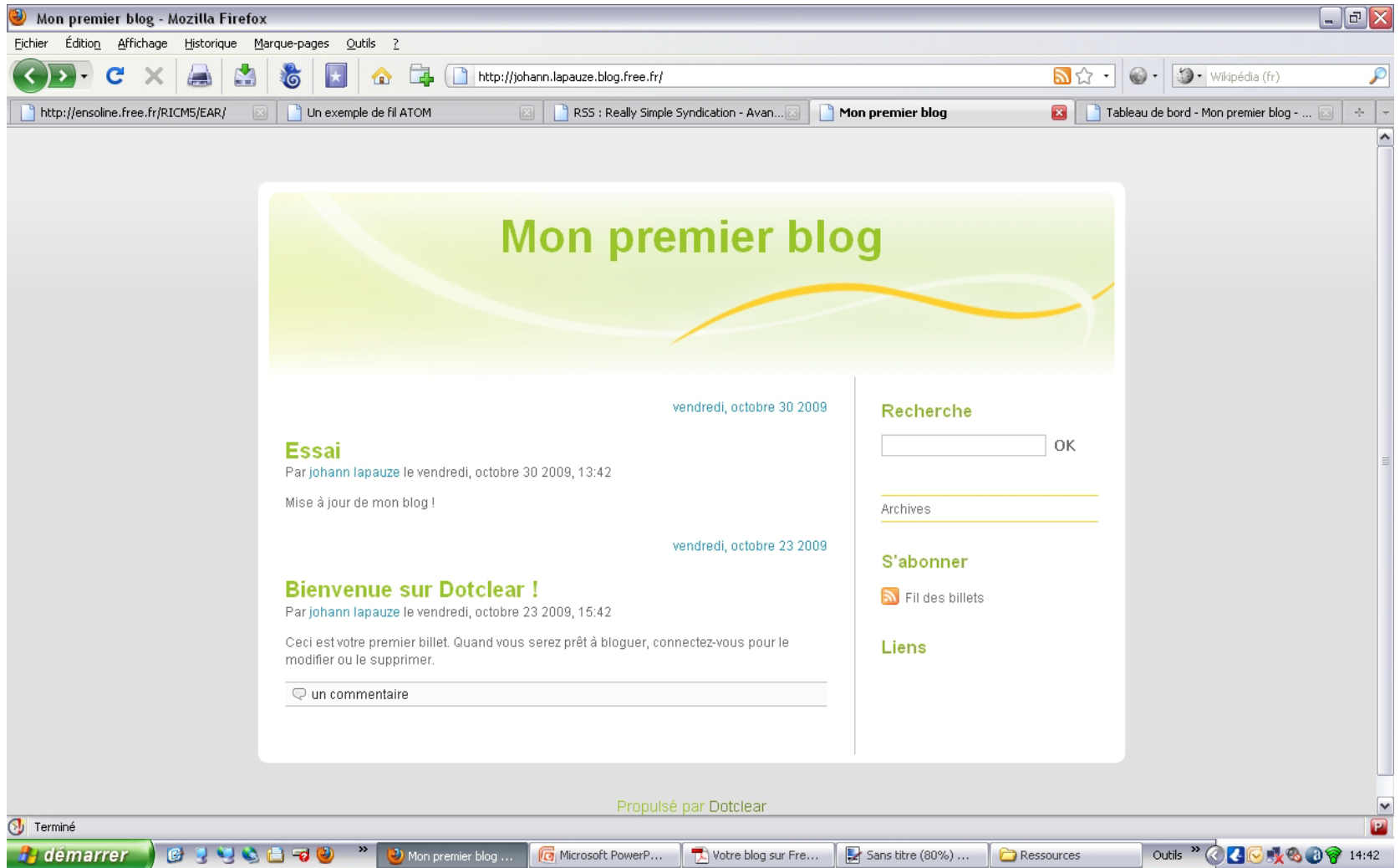
[mise-à-jour sur les chantiers en cours](#)
09 septembre 2009

[Et si on s'associait ?](#)
13 août 2009

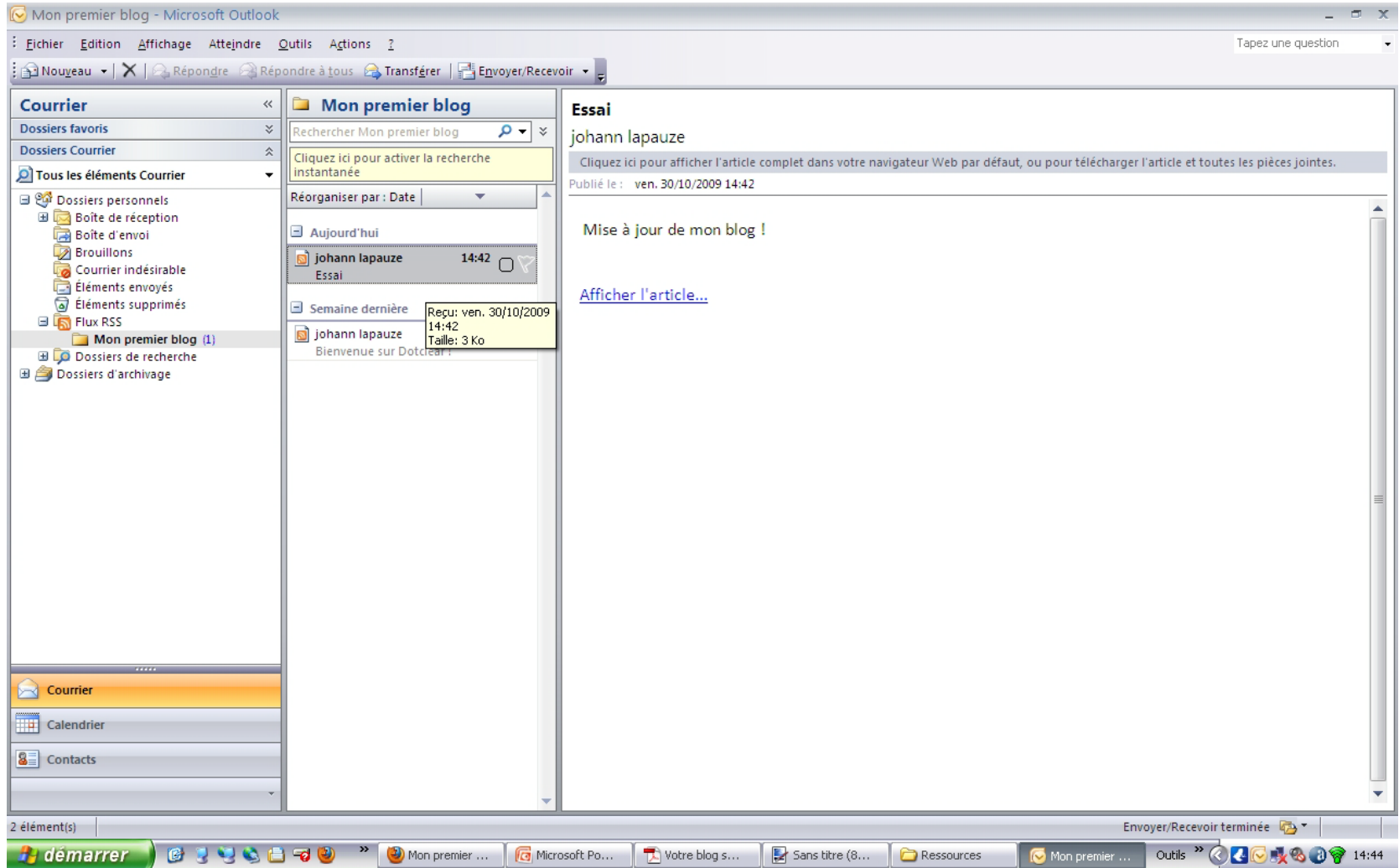
Terminé

démarrer Tableau de bord -... Microsoft PowerP... Votre blog sur Fre... Sans titre - Paint... Ressources Outils 14:42

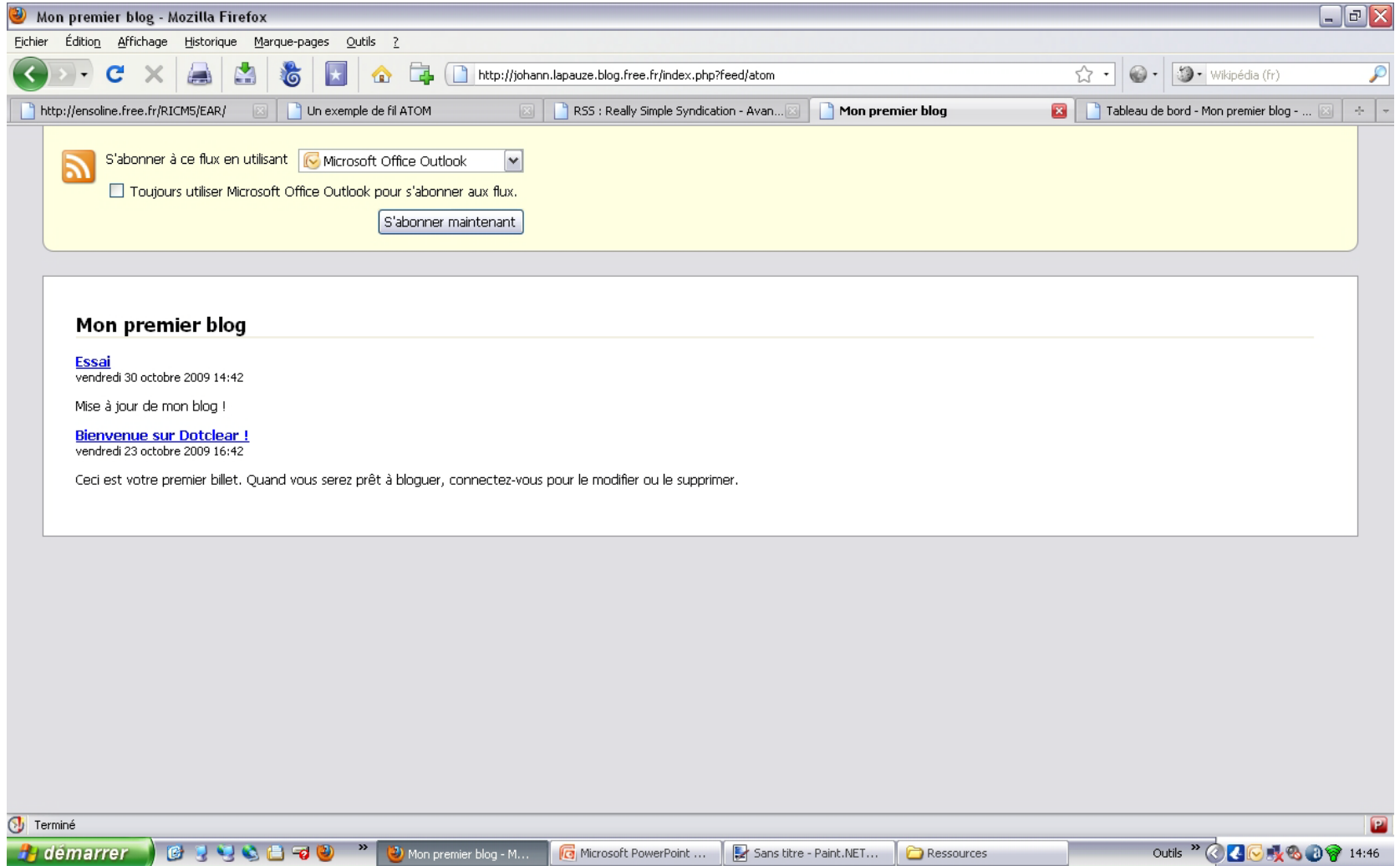
05 - Publication sur le blog



06 - Flux Atom mis à jour sur Outlook



07 - Flux Atom également mis à jour online





4^{ième} Partie – Syndication de contenu

Syndication de contenu

- ▶ Afficher les nouvelles d'un site quelconque sur un autre
- ▶ Ecriture d'un script en perl, php ou Java
 - ▶ Au chargement de la page, exécution du script
 - ▶ Récupération et affichage des informations
- ▶ Annuaire pour lister les flux RSS de divers sites
 - ▶ Syndic8 (<http://www.syndic8.com/>)
 - ▶ Feedster (<http://www.feedster.com/>)

Syndication de contenu

```
<html >
  <head>
    <title>Test RSS</title>
  </head>
  <body>
    <?php
      include("universal-reader.php");
      Universal_Reader("http://192.168.1.1/EaR/demo1/flux.xml");
      echo Universal_Display();
    ?>
  </body>
</html>
```

Syndication de contenu – Librairie tiré de xul

- ▶ Universal_Reader :
 - ▶ Déterminer si c'est un flux RSS ou Atom
 - ▶ Récupérer les entêtes
 - ▶ Lister toutes les entrées du flux

- ▶ Universal_Display :
 - ▶ Construit les balises html pour afficher l'entête
 - ▶ De même pour toutes les entrées du flux

Conclusion

- ▶ Flux RSS et Atom:
 - ▶ Facile à mettre en place
 - ▶ Facile à consulter
 - ▶ Aucune information personnelle à fournir
 - ▶ Possibilité d'insérer du Multimédia

- ▶ Flux RSS vs Atom
 - ▶ Atom est standardisé

- ▶ AtomPub
 - ▶ Peu utilisé aujourd'hui

Perspectives

- ▶ Outils de création de flux RSS pour une page web qui n'en a pas:
 - ▶ Ponyfinish (<http://www.ponyfish.com/>)
 - ▶ Feed43 (<http://www.feed43.com/>)

- ▶ Outil de création de flux RSS même sans site web :
 - ▶ Feedxs (<http://feedxs.com>)
 - ▶ LinkRSS (<http://www.linkrss.com>)

Bibliographie

▶ Techniques :

<http://xml.coverpages.org/rss.html>

<http://cyber.law.harvard.edu/rss/rss.html>

<http://xml.coverpages.org/atom.html>

▶ Comparateur RSS / ATOM :

<http://flurss.free.fr>

<http://www.scriptol.fr/rss/balises-rss-atom.php>

▶ Podcast :

<http://www.apple.com/itunes/podcast/specs.html>

▶ Créer un agrégateur en php:

<http://www.xul.fr/feed/building-feed-reader.html>



5^{ième} Partie – Démonstrations

Environnement

- ▶ Réseau local
 - ▶ Un serveur web WAMP (192.168.1.1)
 - ▶ Un client (192.168.1.2)

- ▶ Architecture du site
 - ▶ /EaR/
 - ▶ demo1/ //Mise en place d'un flux RSS
 - ▶ demo2/ //Mise en place d'un flux Atom
 - ▶ demo3/ //Mise en place de plusieurs flux
 - ▶ demo4/ //Mise en place d'un agrégateur
 - ▶ demo5/ //Mise en place d'un podcast