

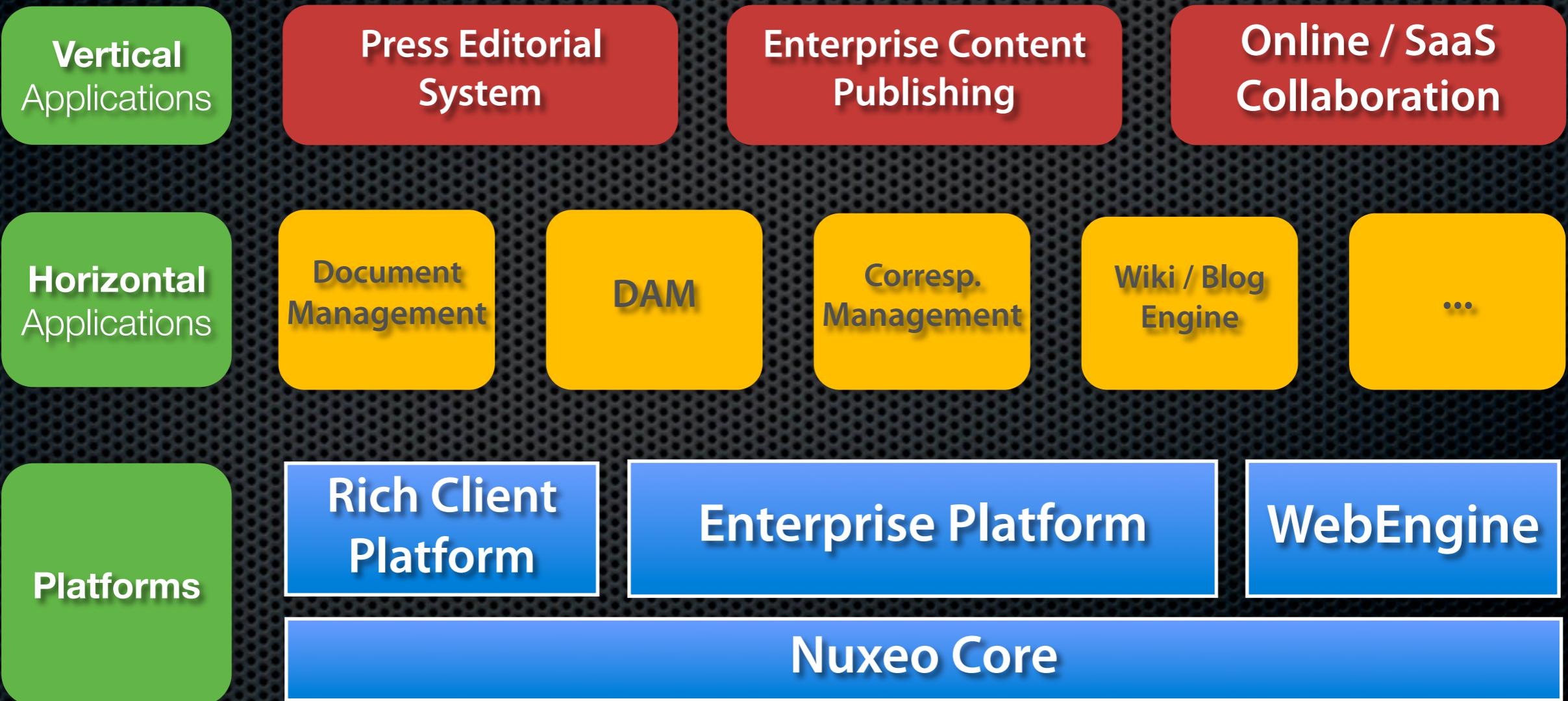
nuxeo • IKS Workshop
Roma

Agenda

- Presentation of the Nuxeo “stack”
 - With a focus on how content is structured and organized (taxonomies, indexing), persistence needs and ongoing work on semantics
- Use cases / demos
 - Default Document Management app
 - RCP apps developed for press agencies

Reminder

Our products: Nuxeo Galaxy



Technology Stack

Nuxeo EP

JavaEE
5

Doc Mgmt / Collab

SI Integration

Enterprise App

Thin Client (Web)

Nuxeo WebEngine

JAX-RS

Social Networking

Wikis / Blogs

Content Publishing

Web Applications

Nuxeo RCP

Eclipse
RCP

Desktop Integration

Business Apps

Productivity Apps

Rich Client

JCR

CMIS

Nuxeo Core

OSGi

Nuxeo Runtime

Java
5

JEE
5



a division of Red Hat



Powered by
jetty://

BUILT ON
eclipse



OpenOffice.org

nuxeo



Content Store Details (1/2)

- Hierarchy of objects (folderish / contentish)
- Formerly based on JCR / Jackrabbit (still supported)
- Now native SQL engine: “Visible Content Storage”
 - Already supports PostgreSQL (prefered), MySQL, Oracle, MS-SQL, H2
 - Uses text indexes native to the RDBMS
- May be extended someday to include a triple or quad store

Content Store Details (1/2)

- Documents are constrained by Document Types
 - Data models are defined by one or more schemas (ex: Dublin Core)
 - Behaviour is expressed by “facets” (ex: Folderish, Indexable, Versionable, etc.)
 - Also room for unconstrained content
- Access rights (ACLs)
- Replications, synchronisation, scalability

Other Persistence Needs

- Workflow: state stored by jBPM
- Relation engine: Jena triple store (currently)
- Directories: store in LDAP or SQL
- Audit: every action in the CMS can be logged to a specific DB for auditing purposes, for social timelines...
- User preferences (personal dashboard, notifications...)
- Conversion: cache for time-expensive transformations (ex: Word -> PDF rendering)

Workflow Details

- Based on jBPM
- Workflows can be graphically designed using jBPM design tools, then seamlessly deployed in Nuxeo
- Used for publication and for more complex document management tasks
- Other WF engines might be plugged with some effort

Searching and Indexing (1/2)

- Used to use Lucene (via Compass) for full text search
 - Didn't prove very reliable (transactionality issues)
- Now text index in the RDBMS
 - Enables mixing fulltext queries with “classical” relational queries (in JOINs...)
 - Performance is DBMS-dependent

Searching and Indexing (2/2)

- Open API for external search engine
 - Enables integration of classical search engine (SOLR) or semantic ones (Sinequa, Exalead, etc.)
 - But not integrated with NXQL / CMISQL
- Tags are usually expressed as “dc:subject” properties
 - Can be controlled or freeform
 - We still need a dedicated service and UI to manage the whole tag space
- Content can be published (w/ WF) in a hierarchical “file plan” as in traditional document management

Relations

- Based on Jena
 - Queries can be expressed in SPARQL
- Current uses:
 - “Depend on”, “Obsoletes”, “Supercedes”... relations between documents
 - Annotations
 - Wiki links (forward / backwards)

Interfaces: Web

- Current “Collaborative DM” interface
 - Based on JSF/Seam
 - Multiple front-ends deployment using standard Java EE architecture (Session Beans)
- WebEngine: lightweight “RESTful” interface
 - Based on the JAX-RS API
 - Easier to customise for mobile devices
 - Used by BBC for mobile site (for instance)

Interfaces: Rich Clients

- RCP
 - Eclipse RCP based, used by AFP and PA
- eRCP
 - Embedded RCP client developed by Yerbabuena Software (community member)
- GWT
 - Used for complex interaction (ex: annotations)
- Flex

Interfaces: Remote API

- Web services:
 - JAX-WS API based on Metro
 - Soon: CMIS WS profile
- REST
 - JAX-RS API (WebEngine)
 - Soon: CMIS AtomPub / REST profile
- Java API: EJB remoting (RMI)

Impacts of Semantics

- Automatic (e)mail classification / routing
- Automatic classification of news, pictures, videos, etc.
- Personalisation of UI
- "Enterprise 2.0" enhancements to collaborative document management (timelines, "friends", etc.)
- Pure "Enterprise 2.0" collaboration suite
 - Wiki + Blog + Forum + Friends
- Modeling (e.g. generating document models from ontologies)

Current Work on Semantics

- Documents (text, images) analysis using UIMA chains
 - Named entities recognition
 - Automatic tagging and classification
 - Scanned documents structure recognition
- Semi-automatic annotations on text and images
 - Ontology extraction from corpora (ex: AFP news archives)
- Experimenting with SKOS for knowledge representation

Inspirations from others

- OpenCalais, Zemanta
- DBPedia, Freebase
- “Subject classification with DITA and SKOS” (on IBM DeveloperWorks)
- “Uncovering the to-dos hidden in your inbox” (in IBM Systems Journal)

Metrics

- 300 KLOC (non-comments)
- 5000 Java classes
 - Including 600 test classes
- 174 modules
- 73 external (open source) libraries

Development tools

- Maven for dependency management
- Mostly Eclipse for development
 - Standard Java / Java EE plugins
 - Specific “Nuxeo IDE” used for faster code->build->test turnaround
 - Work on UML models -> document types, lifecycles and workflow specification generator has recently started

Demos

Demo 1: Nuxeo DM

Non-demo 2: RCP Apps

- AFP “Console NG”
- PA “NU TXT”

AFP Console NG

- Multimedia news authoring and publishing
- Targeted to multimedia Desks
 - XHTML WYSIWYG Editor
 - Full NewsML 1.2
 - IPTC support
 - Real-time collaboration and news feeds
 - Spellchecking support (with custom dictionnaries)
 - Photos / Videos / Flashes
- Java EE on the server side
- Eclipse RCP on the client side

From

AF Consultation par routage FRS

Editor Stockage Outils Bureaux Protocoles Aide

4 ABC84 ETR	EAF.ELU	JPA EMI FRS .0149 070607-08h37 w0354
Pakistan-justice-politique-médias		Crise politique au Pakistan: suspension des mesures de contrôle des médias ISLAMABAD, 7 juin 2007 (AFP) - Le
4 ABC72 ALC	FRS-LC	M EMI FRS .0148 070607-08h36 w0342
Partis-PS-gouvernement-social-fiscalité,LEAD		Hollande: le gouvernement va augmenter les taux de TVA de 2 à 3 points PARIS, 7 juin 2007 (AFP) - Le
4 ABC61 FRA	FAF	PT EMI FRS .0147 070607-08h35 w0342
Partis-PS-gouvernement-social-fiscalité,LEAD		Hollande: le gouvernement va augmenter les taux de TVA de 2 à 3 points PARIS, 7 juin 2007 (AFP) - Le
4 ABC55 FRA	FAF	P EMI FRS .0146 070607-08h34 w0605
Législatives-84,PREU		Dans un climat de peur de l'avenir, le Vaucluse s'ancre à droite (PAPIER D'ANGLE) Par Isabelle WESSELINGH MONTEUX (Vaucluse),
3 ABC11 FRA	FAF	C EMI FRS .0145 070607-08h28 w0248
Histoire-Résistance,LEAD		Décès du général Le Ray, premier chef militaire du maquis du Vercors PARIS, 7 juin 2007 (AFP) - Le général de corps d'armée
4 ABC02 FRA	FAF	P EMI FRS .0144 070607-08h27 w0249
Partis-PS,LEAD		Hollande veut rester premier secrétaire du PS jusqu'au prochain congrès PARIS, 7 juin 2007 (AFP) - François Hollande a indiqué
3 ABB95 ETR	EAF.ELU	JP EMI FRS .0143 070607-08h26 w0292
Indonésie-UE-France-PaysBas-drogue-justice		L'UE demande à l'Indonésie de ne pas exécuter des condamnés à mort JAKARTA, 7 juin 2007 (AFP) - L'Union
4 ABB91 AFA	MOA.PEP	AF EMI FRS .0142 070607-08h25 w0246
GB-Nigeria-littérature-prix		GB: une Nigériane remporte le prix littéraire Féminin Orange Broadband LONDRES, 7 juin 2007 (AFP) - L'écrivain

Feuillet : 0 mot Texte : 0 mot Pas de courrier Paris Phase 3

To

AFP Multimedia Console NG

File Edit Styled Edit Search Window Help

Workspace X

NewsML Metadata

Slug Line: G8-sommet-France-Japon-défense

Head Line: Ventes d'armes à la Chine: le Japon reste opposé à une levée de l'embargo

Sub-Head Line:

Other Sub-Head Lines

Catch Line: Le Premier ministre japonais Shinzo Abe a réaffirmé mercredi l'opposition du Japon à une levée de l'embargo sur les ventes d'armes à la Chine lors d'un entretien avec le Président français Nicolas Sarkozy en marge du sommet du G8.

By Line:

Date Line: HEILIGENDAMM (Allemagne), 6 juin 2007 (AFP) -

Le Premier ministre japonais Shinzo Abe a réaffirmé mercredi l'opposition du Japon à une levée de l'embargo sur les ventes d'armes à la Chine lors d'un entretien avec le Président français Nicolas Sarkozy en marge du sommet du G8.

"Le Japon s'oppose à toute initiative de l'Union européenne en faveur d'une levée de l'embargo sur les exportations d'armes à la Chine", a indiqué M. Abe au Président français, selon un responsable gouvernemental japonais.

"Il est important de connaître la position du Japon sur cette question", a simplement déclaré M. Sarkozy, sans répondre directement au dirigeant nippon.

Feeds Search X

Description	PubDate	Priority	Type	DateLine
Sénat-enquête,2ELEAD Poncelet se dit "victime d'une tentative..." Le président du Sénat Christian...	17:02 06-04	02000C	PARIS (AFI)	
G8-sommet-France Arrivée de Sarkozy à Rostock pour... Le président français Nicolas...	16:48 06-03	POL	ROSTOCK	
Rugby-FRA-NA-TOP-sport	16:22 06-04	SPO	PARIS (AFI)	

Image f... IFO2-512 Animations Video EUA

Description FBL-EURO2008-FINLAND-BEL 21:27 06-06
Finland's Alexei Eremenko (L) celebrates with team-mate Mika Vayrynen after scoring the...

FBL-EURO2008-LIE-ESP 21:27 06-06
Spain's David Villa (C) fires a shot in the middle of Liechtenstein defense (From...)

GERMANY-G8-SUMMIT-ARR: 21:27 06-06
German Chancellor Angela Merkel (L) and her husband Joachim Sauer (2ndL) greet...

GERMANY-G8-SUMMIT-ARR: 21:26 06-06
German Chancellor Angela Merkel (L) and her husband Joachim Sauer (2nd L)...

060607DUS251 21:26 06-06
Fussball, Laenderspiel, Qualifikation EM 2008, Deutschland - Slowakei,...

GERMANY-G8-SUMMIT-ARR: 21:26 06-06
German Chancellor Angela Merkel (L) and her husband Joachim Sauer (2ndL) greet...

Quick View X

GERMANY-G8-SUMMIT-ARRIVAL-LUCKOW

SubHeadLines

No sub head lines

GERMANY-G8-SUMMIT-ARRIVAL-LUCKOW

German Chancellor Angela Merkel (L) and her husband Joachim Sauer (2ndL) greet Italian Prime minister Romano Prodi and his wife Flavia Prodi (R) upon their

Feeds Search size:56; page:1 / 1

User:doudet

Role:Advanced Editor

45M of 72M

The Workspace

Collaborative Workspace

Publishing Sections

Metadata & Rich Text Editor

Search Results

Incoming Feeds

Quick View

This screenshot shows the AFP Multimedia Console NG workspace, which is a collaborative environment for managing news content. The interface is divided into several panels:

- Left Panel:** A tree view of the "workspace" structure, showing various news items and sections like "Top 14/XV de France - Blanco: pas de tour", "Le cinéma", "Interdiction", "Le journal vidéo de l'AP", "Roland-Garros", "10 à 12 ans de prison requis aux assises", "NFL Europe", "Sommet Merkel-Blair", "Montadori rachète EMAP", "Affaire Litvinenko", "Hollande chiffre", "CORSE", "Pakistan", and "FRANCE-MEDIA-LEMONDE". Below it is a "Rubrics" panel.
- Central Panel:** The main workspace for editing news items. It includes fields for "Slug Line" (G8-sommet-France-Japon-défense), "Head Line" (Ventes d'armes à la Chine: le Japon reste opposé à une levée de l'embargo), "Sub-Head Line", "Other Sub-Head Lines", "Catch Line" (Le Premier ministre japonais Shinzo Abe a réaffirmé mercredi l'opposition du Japon à une levée de l'embargo sur les ventes d'armes à la Chine lors d'un entretien avec le Président français Nicolas Sarkozy en marge du sommet du G8.), "By Line", "Date Line" (HEILIGENDAMM (Allemagne), 6 juin 2007 (AFP) -), and a rich text editor containing the full article text. The rich text editor is highlighted with a yellow box.
- Bottom Left Panel:** A "Feeds Search" panel showing search results for "Sénat-enquête,2ELEAD" (PubDate: 17:02 06-04, Type: 02000 PARIS (AFI)), "G8-sommet-France" (PubDate: 16:48 06-03, Type: POL ROSTOCK), and "Rugby-FRA-NA-TOP-sport" (PubDate: 16:22 06-04, Type: SPO PARIS (AFI)).
- Bottom Right Panel:** A "Quick View" panel for the "GERMANY-G8-SUMMIT-ARRIVAL-LUCKOW" feed, showing a photo of Angela Merkel, Joachim Sauer, and Romano Prodi greeting each other.
- Right Panel:** A "PubDal" panel displaying a list of incoming feeds, including "FBL-EURO2008-FINLAND-BEL", "FBL-EURO2008-LIE-ESP", "GERMANY-G8-SUMMIT-ARR", "GERMANY-G8-SUMMIT-ARR", "060607DUS251", and "GERMANY-G8-SUMMIT-ARR". Each item has a thumbnail image and a brief description.

Workflow

1 Incoming Feeds

2 Metadata & Rich Text Editor

3 Collaborative Workspace

4 Publishing Sections

1b Search Results

The screenshot displays the AFP Multimedia Console NG application window with several panels:

- Left Panel:** Collaborative Workspace (3) and Publishing Sections (4). The Publishing Sections panel shows a tree view of news categories like La Une, France, Internationale, Afrique, etc.
- Middle Panel:** NewsML Metadata and Rich Text Editor (2). It contains fields for Slug Line, Head Line, Sub-Head Line, Catch Line, and Date Line. The Rich Text Editor area contains two paragraphs of text about the G8 summit and Japanese opposition to arms exports to China.
- Bottom Panel:** Feeds Search (1b) showing search results for "Sénat-enquête, 2ELEAD" and "G8-sommet-France".
- Right Panels:** Incoming Feeds (1), Quick View, and SubHeadLines. The Incoming Feeds panel lists various news items with thumbnails and details. The Quick View panel shows a preview of a German-Chancellor-Angela-Merkel photo. The SubHeadLines panel shows "No sub head lines".

It's even bigger, actually



Rich Text Editor

AFP Multimedia Console NG

File Edit Styled Edit Search Window Help

Paragraph | A A U | [bold] [italic] [underline] | [list] [list] [list] | [list] [list] | [list] [list] | [list] [list] | [list] [list]

Fifa/Altitude-Bolivie: Evo Morales se rendra à la réunion de ... | Ventes d'armes à la Chine: le Japon reste opposé à une levé... | *Arrivée de Sarkozy à Rostock pour le sommet du G8

NewsML Metadata

Slug Line: G8-sommet-France

Head Line: Arrivée de Sarkozy à Rostock pour le sommet du G8

Sub-Head Line:

▶ Other Sub-Head Lines Reset Sub-Head Lines

Catch Line: Le président français Nicolas Sarkozy est arrivé mercredi après-midi à l'aéroport de Rostock-Laage (nord-est de l'Allemagne), pour participer au sommet du G8 qui se tient dans la station balnéaire de Heiligendamm, du 6 au 8 juin.

By Line:

Date Line: ROSTOCK (AFP) -

Credit Line:

Language: fr

Source: AFP

Type: CLT CLJ DIS ECO EDU
 ENV HTH HUM SOC LIF
 POL REL SCI SOI SPO



Le président français Nicolas Sarkozy est arrivé mercredi après-midi à l'aéroport de Rostock-Laage (nord-est de l'Allemagne), pour participer au sommet du G8 qui se tient dans la station balnéaire de Heiligendamm, du 6 au 8 juin.

L'Airbus présidentiel, en provenance de Paris, s'est posé peu avant 14h00 GMT, selon les images transmises par la télévision interne du G8.

M. Sarkozy est descendu de l'avion main dans la main avec son épouse Cécilia, toute vêtue de blanc. Le couple devait immédiatement se rendre en hélicoptère à Heiligendamm où il réside, comme les autres chefs de délégation, dans un luxueux complexe hôtelier au bord de la Baltique.

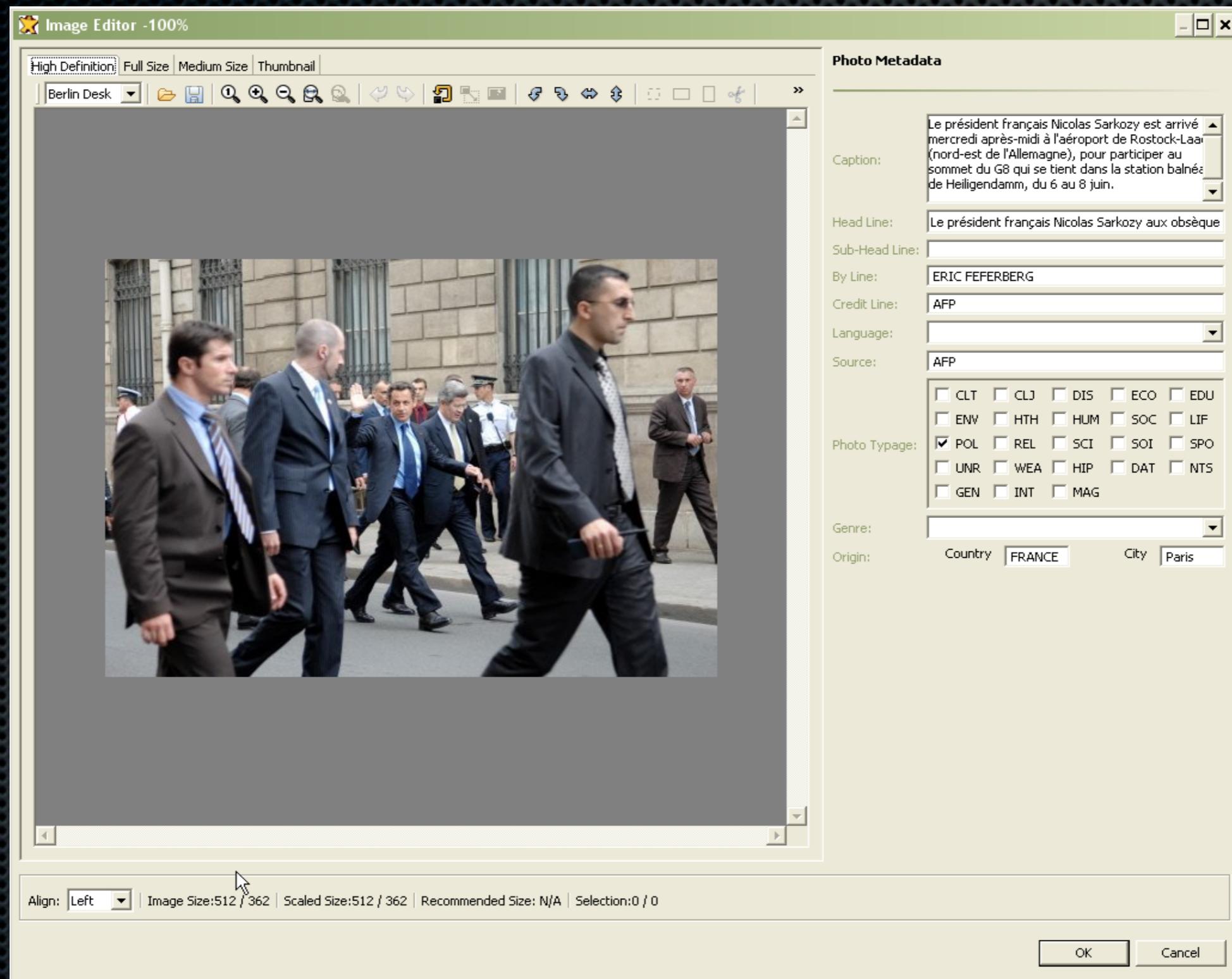
Le président français, qui a succédé le 16 mai au vétéran Jacques Chirac, participe à son premier sommet des huit grandes puissances.

Il devait aussitôt s'entretenir avec la chancelière allemande Angela Merkel, puis avec le Premier ministre japonais Shinzo Abe.

Edit Preview

Current element:P P:5 W:154 C:866 W:41 / C:217 User:doudet Role 47M of 72M

Flexible Picture Editor



Delivered to customers

Yahoo! My Yahoo! Mail

YAHOO! NEWS [Sign In](#)
New User? Sign Up

Search the Web Search

News Home - Help

[News Home](#) [U.S.](#) [Business](#) [World](#) [Entertainment](#) [Sports](#) [Tech](#) [Politics](#) [Science](#) [Health](#) [Most Popular](#) [Index](#)

[Photos](#) | [Opinion](#) | [Local News](#) | [Odd News](#) | [Comics](#) | [Weather](#) | [Full Coverage](#) | [Video/Audio](#) | [In the Hot Zone](#)

News Search Now with Blogs (BETA) All News & Blogs Advanced

Top Stories - AFP



France in crisis as riots spread

2 hours, 57 minutes ago
PARIS (AFP) - Prime Minister Dominique de Villepin and Interior Minister Nicolas Sarkozy held a crisis meeting on running riots that have plunged France into its worst explosion of urban violence in decades.

View: [Headlines Only](#) | [Include Summaries](#) | [Include Photos](#)



Thousands oppose Bush at Americas summit

2 hours, 2 minutes ago
MAR DEL PLATA, Argentina (AFP) - Thousands staged angry anti-US protests as President George W. Bush sought to convince the 34-nation Summit of the Americas to follow a free trade agenda.



Pakistan's Musharraf criticises West for quake aid shortfall

Fri Nov 4, 10:28 AM ET
MUZAFFARABAD, Pakistan (AFP) - Pakistani President Pervez Musharraf accused the West of double standards for failing to meet the country's quake aid needs, adding that he would delay buying US fighter jets to focus on relief efforts.



Iraq's PM uses Muslim holiday to urge insurgency end

2 hours, 35 minutes ago
BAGHDAD (AFP) - At least 15 Iraqis were killed in separate incidents across the country as Prime Minister Ibrahim Jaafari used the occasion of the Eid al-Fitr holiday to urge insurgents to abandon their fight.

ADVERTISEMENT

Ecoutez la radio !

Choisissez une station :

 Top Pop  Rock  Fan des années 80  Love FM  Rap  RnB  Pop/Rock français  Numéros 1  Jazz  Divas & Crooners

 [Toutes les stations >](#)

MORE IN YAHOO! NEWS

- AP
- Reuters
- **AFP**
- USATODAY.com
- CSMonitor.com
- Los Angeles Times
- Chicago Tribune
- U.S. News & World Report
- NPR
- Reuters Features

TOP STORIES SLIDESHOES

OCTOBER PAYROLLS WEAKER THAN EXPECTED
5.0% 

Hurricane Katrina

nuxeo

NUTXT Project

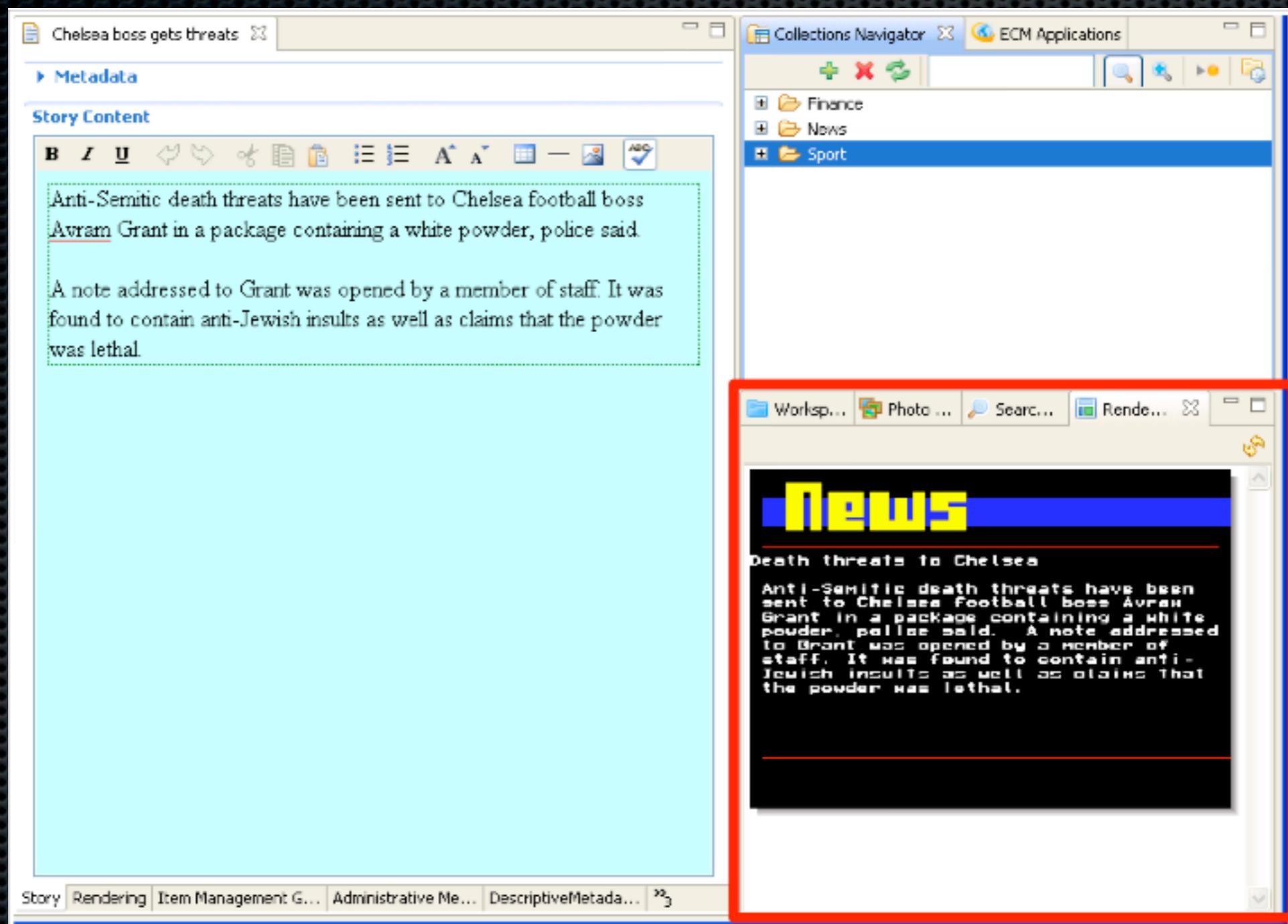
- New editorial system for PA to serve analog and digital Teletext in the UK
- Create the foundation for other systems at PA that might migrate to it (photo, text)
- 100+ users, 4 locations (UK, India)

Workspace

The screenshot displays the Nuxeo workspace interface with four main panels:

- Collaborative "Baskets" (Top Left):** A grid view of documents under the "SPORT" basket, filtered by "Last 7 Days". A yellow box highlights the "Baskets" tab in the top navigation bar.
- Metadata and Content Editor (Top Middle):** An editor for the document "Chelsea boss gets death threats". It shows fields for Title, Author, Topic, Region, Creation Time, Modification Time, Last User, Embargo Date, Expiry Date, and Required update time. A yellow box highlights the "Metadata" section.
- Collection / Slots Management (Top Right):** A tree view of collections under "Sport". The "Cricket" collection is expanded, showing "Football", "Golf", and "Tennis" slots. A yellow box highlights the "Collection / Slots Management" section.
- Search Photo Collection etc. (Bottom Right):** A search results table for the "Cricket_Story_Template" and "Ponting worried" documents. A yellow box highlights the search results area.

Live Preview



Custom Dictionnaries

Story Content

B I U ⌛ ⌚ ⌚ ⌚ A A ABC

India skipper Mahendra Dhoni attracted the highest early price as bidding got under way for players in the Indian Premier League.

Some star-studded names are tournamant.

Dhoni was bought by Chennai (£359,410) and Shane Warne

The auction continues throughout the day in Mumbai on Wednesday.

Spelling Suggestions ▾
Spelling Dictionaries ▾

Undo Ctrl+Z
Cut Ctrl+X
Copy Ctrl+C
Paste Ctrl+V

Add To Dictionary ▾
English (GB)
Cricket