

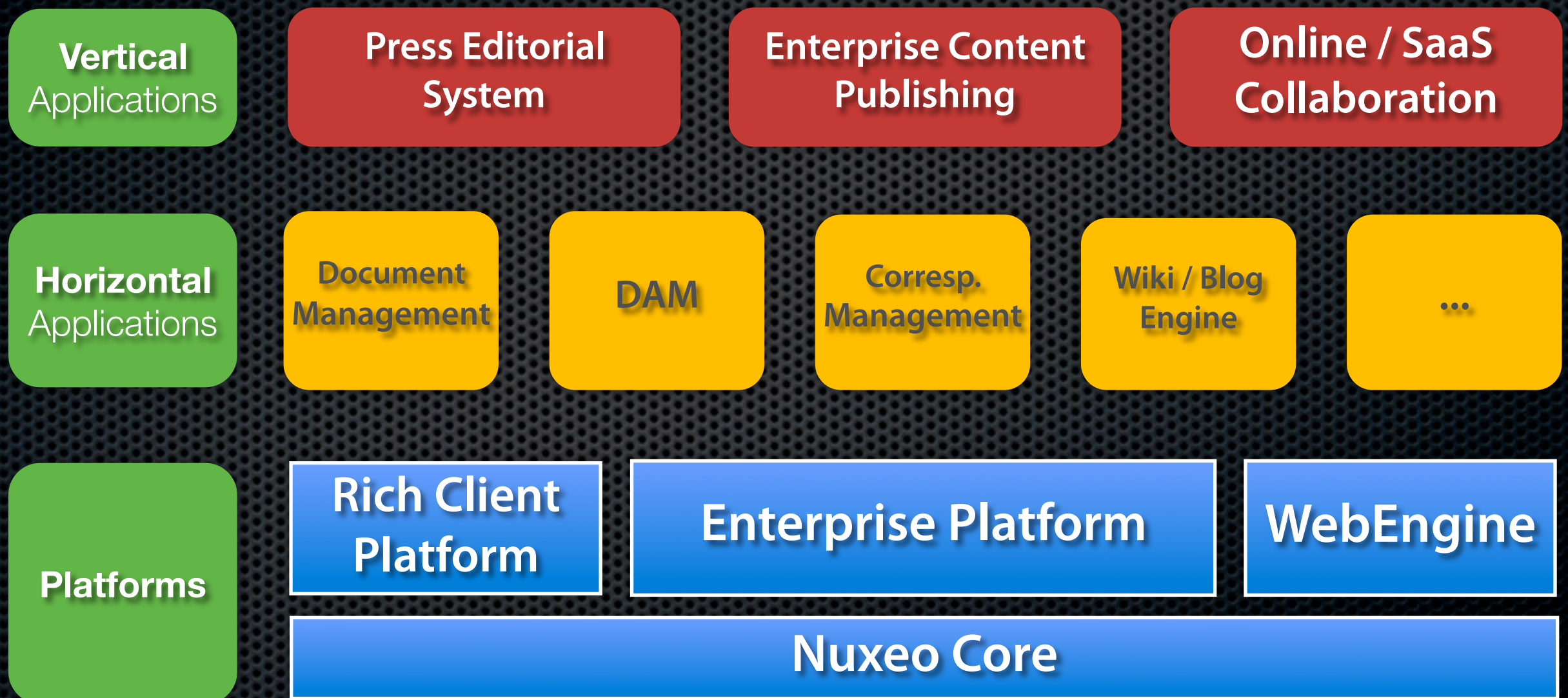
nuxeo ● **IKS Workshop Roma**

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

Nuxeo Runtime

OSGi

Java
5



JEE
5



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 (preferred), 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 dictionnaires)
 - ✦ Photos / Videos / Flashes
- ✦ Java EE on the server side
- ✦ Eclipse RCP on the client side

From

AFP Consultation par routage FRS									
Editer 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	w0203
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					

To

AFP Multimedia Console NG

File Edit Styled Edit Search Window Help

Workspace

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

Reset 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.

Edit Preview

Feeds Search







Description	PubDate	Priority	Type	DateLine
 Sénat-enquête, 2E LEAD Poncelet se dit "victime d'une tentative de..." Le président du Sénat Christian...	17:02 06-04	4	020000	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)

Feeds Search size:56; page:1 / 1

User:doudet Role:Advanced Editor 45M of 72M

Image f... IFO2-512 Animations Video EUA

Description PubDal


	FBL-EURO2008-FINLAND-BEL Finland's Alexei Eremenko (L) celebrates with team-mate Mika Vayrynen after scoring the...	21:27 06-06
	FBL-EURO2008-LIE-ESP Spain's David Villa (C) fires a shoot in the middle of Liechtenstein defense (From...	21:27 06-06
	GERMANY-G8-SUMMIT-ARR German Chancellor Angela Merkel (L) and her husband Joachim Sauer (2ndL) greet...	21:27 06-06
	GERMANY-G8-SUMMIT-ARR German Chancellor Angela Merkel (L) and her husband Joachim Sauer (2nd L)...	21:26 06-06
	060607DUS251 Fussball, Laenderspiel, Qualifikation EM 2008, Deutschland - Slowakei,...	21:26 06-06
	GERMANY-G8-SUMMIT-ARR German Chancellor Angela Merkel (L) and her husband Joachim Sauer (2ndL) greet...	21:26 06-06

Quick View

GERMANY-G8-SUMMIT-ARRIVAL-LUCKOW

SubHeadLines

No sub head lines



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 arrival at the airport in Luckow.

The Workspace

The screenshot displays the AFP Multimedia Console NG workspace, a complex interface for managing news content. The interface is divided into several main sections:

- Left Panel (Collaborative Workspace):** A list of news items with titles such as "Poncelet se dit 'victime d'une tentative de..." and "Arrivée de Sarkozy à Rostock pour le somm...". A yellow callout box labeled "Collaborative Workspace" is overlaid on this panel.
- Top Panel (NewsML Metadata):** A form for editing metadata for a selected item. It includes fields for Slug Line, Head Line, Sub-Head Line, Catch Line, By Line, and Date Line. A yellow callout box labeled "Metadata & Rich Text Editor" is overlaid on this panel.
- Right Panel (Incoming Feeds):** A list of incoming news feeds with titles like "FBL-EURO2008-FINLAND-BEL" and "GERMANY-G8-SUMMIT-ARR...". A yellow callout box labeled "Incoming Feeds" is overlaid on this panel.
- Bottom Left Panel (Publishing Sections):** A list of publishing sections including "La Une", "France", "Politique", "Monde", "Afrique", "Economie/Finances", "Sport", "Football", "Culture/Art de vivre", "Média", "Sciences/Environnement", "Les gens", "Insolite", "High Tech", "Medecine/santé", "Titre de la rubrique Evènement", "Evènement", and "Documentation". A yellow callout box labeled "Publishing Sections" is overlaid on this panel.
- Bottom Center Panel (Search Results):** A table showing search results with columns for Description, PubDate, Priority, Type, and DateLine. The results include items like "Sénat-enquête, 2LEAD" and "G8-sommet-France". A yellow callout box labeled "Search Results" is overlaid on this panel.
- Bottom Right Panel (Quick View):** A detailed view of a selected item, showing a title "GERMANY-G8-SUMMIT-ARRIVAL-LUCKOW", a sub-headline, and a description. A yellow callout box labeled "Quick View" is overlaid on this panel.

The interface also includes a top menu bar with options like File, Edit, Styled Edit, Search, Window, and Help, and a bottom status bar showing "Feeds Search size:56; page:1 / 1", "User:doudet", "Role:Advanced Editor", and "45M of 72M".

Workflow

AFP Multimedia Console NG

File Edit Styled Edit Search Window Help

Collaborative Workspace (3)

NewsML Metadata (2)

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: HEILIGENDAMM (Allemagne) 17 juin 2007 (AFP) -

Date Line: 17:02 06-06

Metadata & Rich Text Editor

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.

Incoming Feeds (1)

Publishing Sections (4)

Search Results (1b)

Quick View

Feeds Search

Description	PubDate	Priority	Type	DateLine
Sénat-enquête, 2E LEAD Poncelet se dit "victime d'une tentative de..." Le président du Sénat Christian...	17:02 06-06	4	020000	PARIS (AFI)
G8-sommet-France Arrivée de Sarkozy à Rostock Le président français Nicolas...	16:48 06-06	3	POL	ROSTOCK
Rugby-FRA-NA-TOP-sport	16:22 06-06	4	SPO	PARIS (AFI)

Feeds Search size:56; page:1 / 1

User:doudet Role:Advanced Editor 45M of 72M












It's even bigger, actually




































Rich Text Editor

AFP Multimedia Console NG

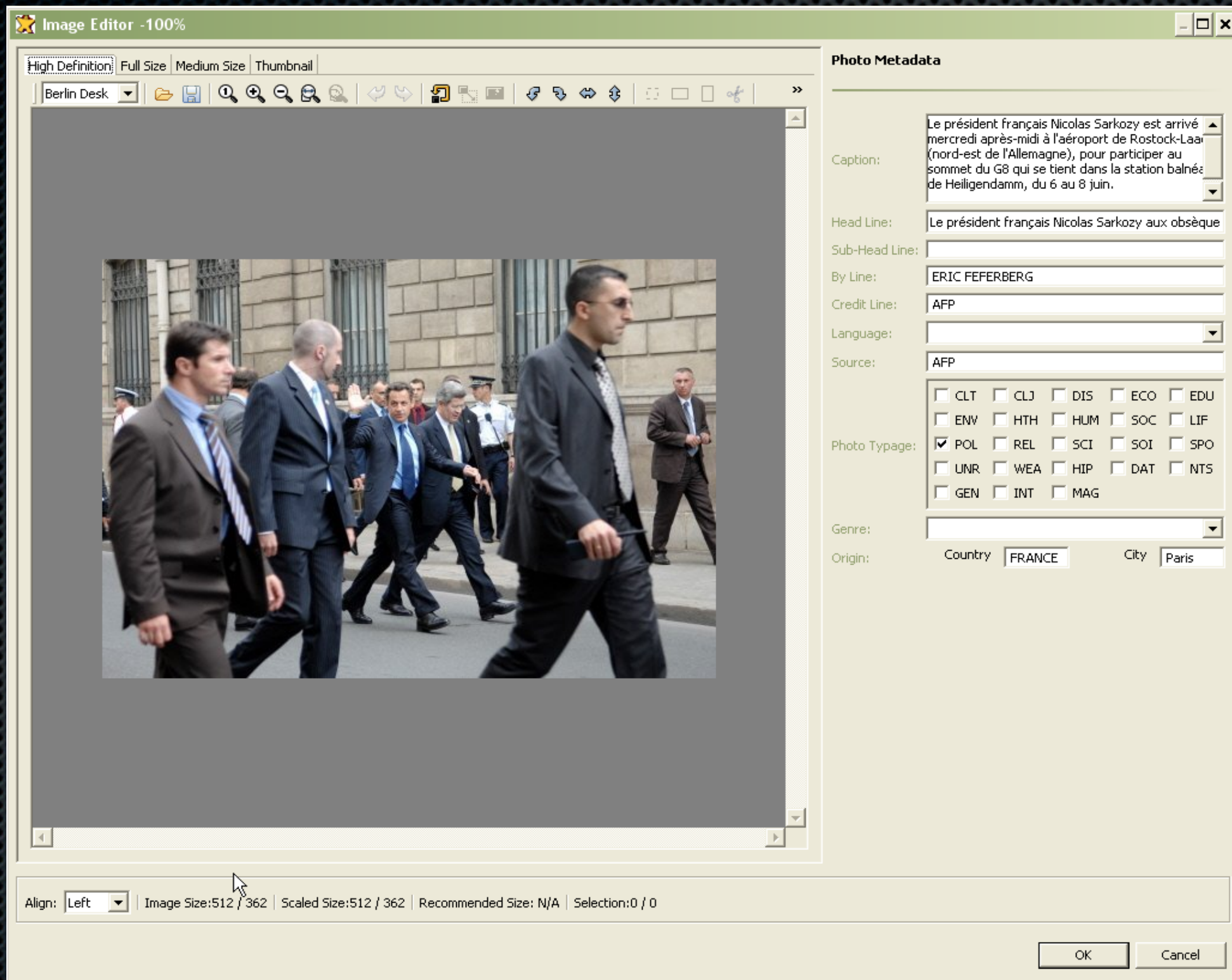
File Edit Styled Edit Search Window Help



Paragraph



Flexible Picture Editor



Delivered to customers

Yahoo! My Yahoo! Mail

Search the Web Search

YAHOO! NEWS Sign In New User? Sign Up 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 (GETA) ch: All News & Blogs Search 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: Alanis Morissette, Anastacia, Air, et plus encore [Ecoutez !](#)
Rock: Ash, B.B. King, Belle & Sebastian et plus encore [Ecoutez !](#)
Fan des années 80: [Pop/Rock français](#)
Love FM: [Numéros 1](#)
Rap: [Jazz](#)
RnB: [Divas & Crooners](#)
[MUSIC radio](#) [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 SLIDESHOWS

OCTOBER PAYROLLS WEAKER THAN EXPECTED [Hurricane Katrina](#)

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

PressAssociation NuTxt

File Window Security Doc ops Help Areas

Baskets

SPORT Chelsea legend

No state filtering Last 7 Days

Title	Last mod...	Last user	Embargo	Expire
Mahendra Dhoni bidding	20 févr. ...	Administ...		21 févr. 2008 13:38
Chelsea boss gets de...	20 févr. ...	Administ...		
Ponting worried	20 févr. ...	Administ...	21 févr. 2008...	22 févr. 2008 13:50
Barca legend	20 févr. ...	Administ...	22 févr. 2008...	23 févr. 2008 13:17

Collaborative "Baskets"

Chelsea boss gets death threats

Metadata

Title: Chelsea boss gets death threats

Author: Administrator

Topic:

Region:

Creation Time: mercredi 20 février 2008 11:27:11

Modification Time: mercredi 20 février 2008 11:28:58

Last User: Administrator

Embargo Date: mercredi 20 février 2008 14:14:15

Expiry Date: mercredi 20 février 2008 14:14:15

Required update time: mercredi 20 février 2008 14:14:15

Metadata and Content Editor

Brief Index Text:

Short Index Text:

Medium Index Text:

Long Index Text:

URL:

Trail:

Headline:

Story Content

Collections Navigator

ECM Applications

- Finance
- News
- Sport
 - Cricket
 - Football
 - Golf
 - Tennis

Collection / Slots Management

Collection Editor : Cricket

Index Editor : Index_Cricket

Slot	Stage	Air
00		
01		
02		
03		
04		
05		
06		
07		

Publishing area Staging / On Air

Workspace Navi...

Photo Basket Search Result

Title	Last User	Basket
Cricket_Story_Template	Administrator	Sport
Ponting worried	Administrator	SPORT

Search Photo Collection etc.

Current mode : View Mode.

Story Rendering Item Management G... Administrative Me... DescriptiveMetada...

Live Preview

The screenshot displays a software interface for document management and rendering. The main window is titled "Chelsea boss gets threats" and contains a "Metadata" section and a "Story Content" section. The "Story Content" section has a rich text editor with a toolbar and a large text area containing the following text:

Anti-Semitic death threats have been sent to Chelsea football boss Avram Grant in a package containing a white powder, police said.

A note addressed to Grant was opened by a member of staff. It was found to contain anti-Jewish insults as well as claims that the powder was lethal.

To the right of the main window is a "Collections Navigator" panel with tabs for "Finance", "News", and "Sport". The "News" tab is selected.

Below the main window is a "Worksp..." panel with tabs for "Photo ...", "Search...", and "Rende...". The "Rende..." tab is selected, showing a rendered output of the document. The rendered output features a yellow "News" header and the same text as the main window, but in a black and white, high-contrast format.

The bottom of the interface shows a status bar with tabs for "Story", "Rendering", "Item Management G...", "Administrative Me...", and "DescriptiveMetada...".

Custom Dictionaries

Story Content

B I U [undo] [redo] [cut] [copy] [paste] [bulleted list] [numbered list] [font size up] [font size down] [background color] [text color] [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 (\$1,000,000) and Australia's Adam Gilchrist went for \$700,000.

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

Dhoti