



EBU Open source seminar
Geneva 1-2 October

Where to use open source

Giorgio Dimino
RAI – Research Centre
g.dimino@rai.it



Archive ingestion projects

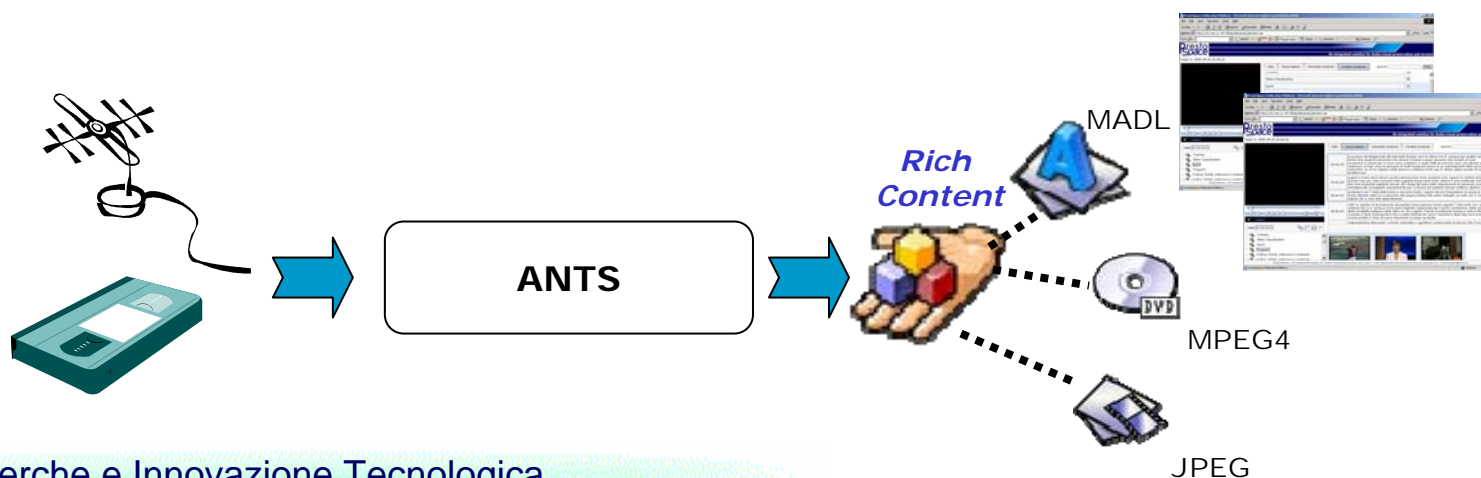
- ✓ RAI Research Centre has been involved for many years in developing ingestion system components for the archive department
- ✓ Main projects
 - ☞ Tv and Radio emission logging
 - ☞ Automated robotic digitisation of Betacam tapes
 - ☞ Automatic metadata extraction
 - ☞ 1/4" audio tapes digitisation
- ✓ All of these projects make large use of free software



ANTS

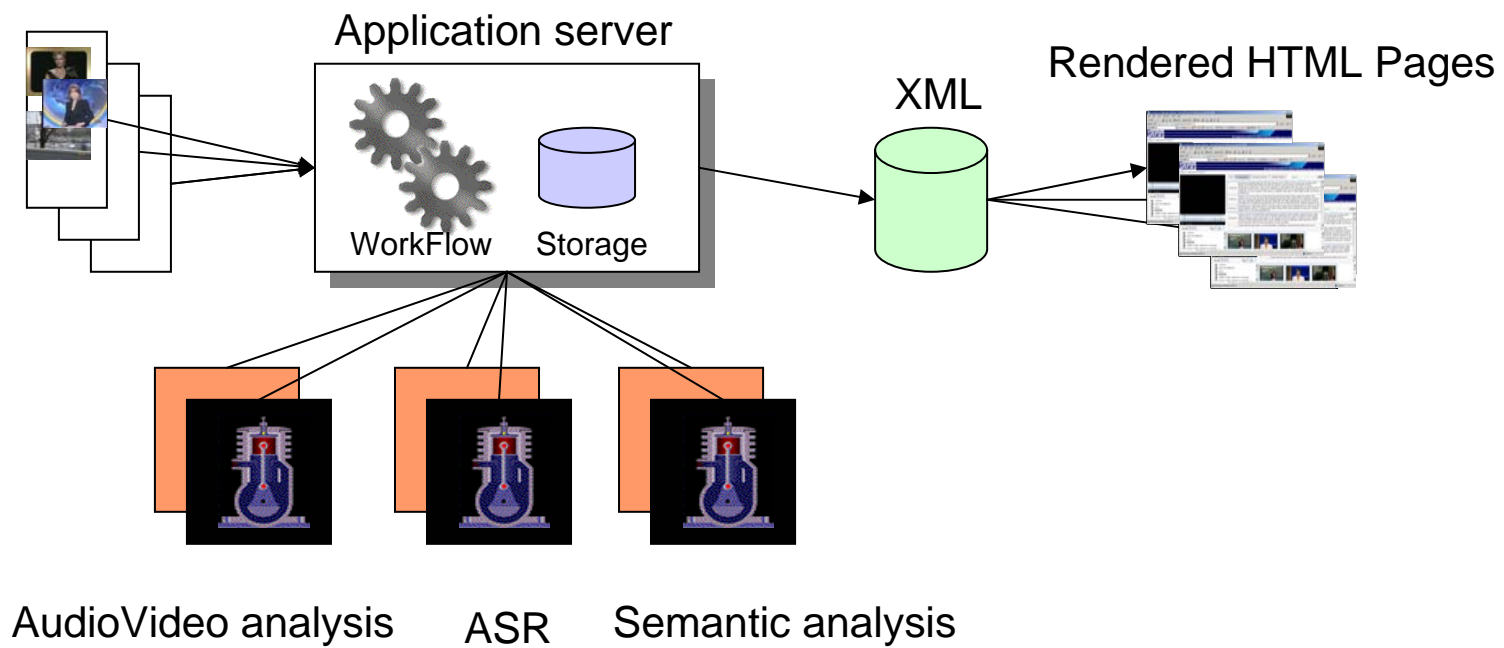
Automatic Newscast Transcription System

- ✓ The ANTS project aims to reduce the documentation costs and timing via automatic analysis of the audiovisual content
- ✓ Main target is the news environment
- ✓ The project started in 2001
- ✓ Two production releases have been made available up to now

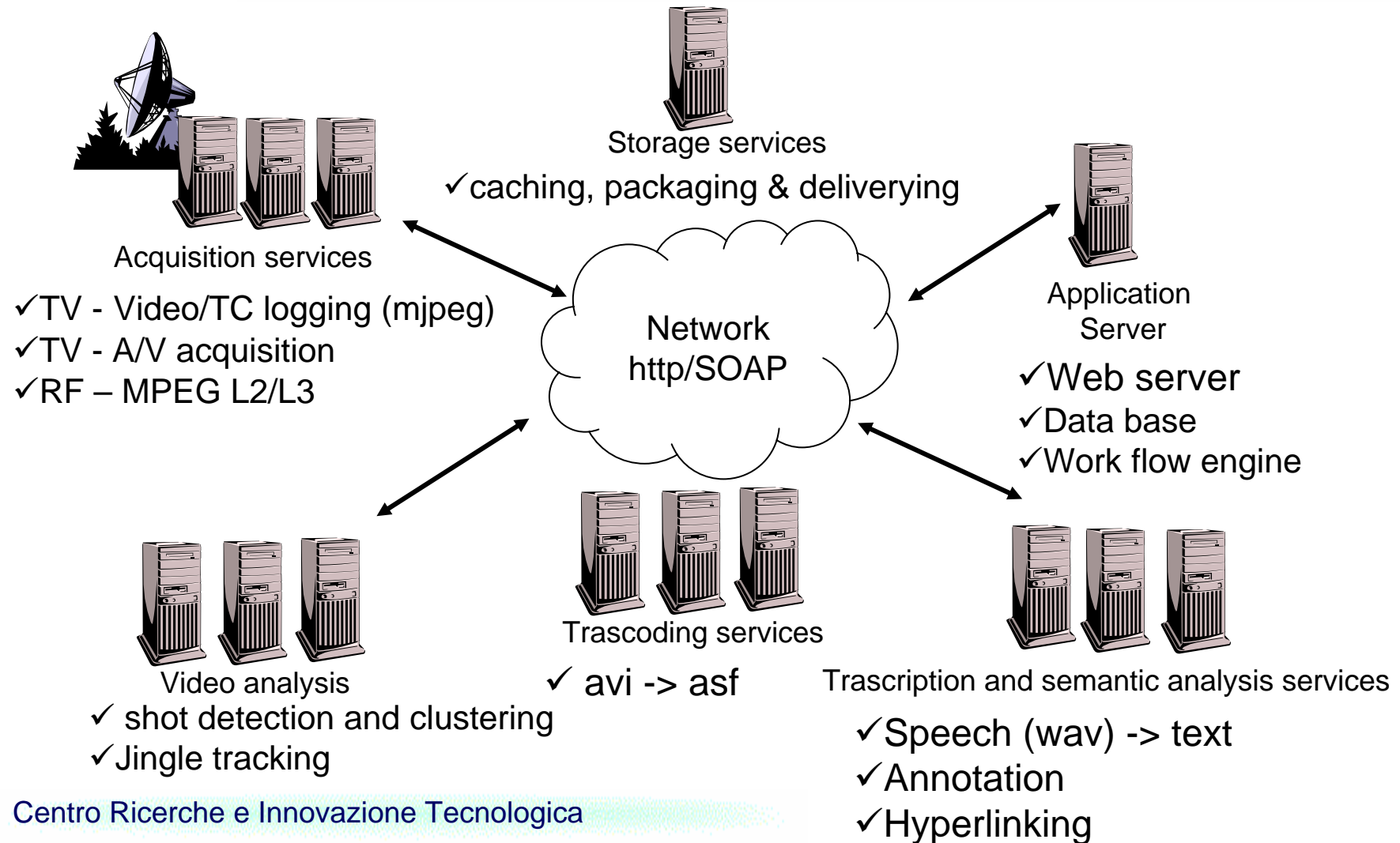


ANTS Architecture

AV Capture Stations



ANTS sytem overview





ANTS WorkFlow Manager

Home

ANTS

My Profile

User logout

Consulting

Octopus

AntsSearch

Workflow

Summary Report

Active Report

Fallout Report

Inactive Report

Mixed Report

Process Graph

Client Control

Acquisition

Scheduler

Multimedia Cut

Programme Cut

Cut Agenda

Channel Settings

New Scheduling

Prog. Management

Administration

Add New User

Manage Users

Add/Rem Hosts

TRS Groups

Configuration

Send Notify

Message List

Dictionary Editor

Help

Documentation

Today is : 2006-03-07 14:37:07

You are logged in as admin - Product Version 2.0b

<input checked="" type="checkbox"/>	RAI3	TGR Piemonte	2006-3-05 14:00:02	2006-3-05 14:31:21	ANTS_1141737397.22	95.9%	active	<input type="checkbox"/>
<input checked="" type="checkbox"/>	RAI1	TG1	2006-3-05 13:29:50	2006-3-05 13:56:35	ANTS_1141737012.61	100.0%	active	<input type="checkbox"/>
<input checked="" type="checkbox"/>	CANALE5	TG5	2006-3-05 13:00:00	2006-3-05 13:39:00	none	100.0%	terminated	<input type="checkbox"/>
<input checked="" type="checkbox"/>	RAI2	TG2	2006-3-05 13:00:00	2006-3-05 13:30:00	none	100.0%	terminated	<input type="checkbox"/>
<input checked="" type="checkbox"/>	RAI2	TG2	2006-3-04 20:30:00	2006-3-04 21:00:00	none	100.0%	terminated	<input type="checkbox"/>
<input checked="" type="checkbox"/>	RAI1	TG1	2006-3-04 20:00:00	2006-3-04 20:30:00	none	100.0%	terminated	<input type="checkbox"/>
<input checked="" type="checkbox"/>	CANALE5	TG5	2006-3-04 20:00:00	2006-3-04 20:30:00	none	100.0%	terminated	<input type="checkbox"/>
<input checked="" type="checkbox"/>	RAI3	TGR Piemonte	2006-3-04 19:34:43	2006-3-04 19:53:24	ANTS_1141661778.05	100.0%	complete	<input type="checkbox"/>
<input checked="" type="checkbox"/>	RAI3	TG3	2006-3-04 19:00:22	2006-3-04 19:30:33	ANTS_1141661482.93	100.0%	complete	<input type="checkbox"/>
<input checked="" type="checkbox"/>	RAI3	TG3	2006-3-04 14:23:28	2006-3-04 14:51:06	ANTS_1141661385.96	23.8%	complete	<input type="checkbox"/>
Program Details								
Selected Option								
Speech To Text		yes	ANTS_1141661385.96	7296	RAI3_TG3_1430_daily.xml	User : admin Date : 2006-3-06 Time : 17:09:01	rsv	Multimedia Info S.Date : 2006-03-04 S.Time : 14:23:28.62 E.Date : 2006-03-04 E.Time : 14:30:02.05 Host : IISTOAQ2 S.Date : 2006-03-04 S.Time : 14:30:02.09 E.Date : 2006-03-04 E.Time : 14:30:03.16 Host : IISTOAQ1 S.Date : 2006-03-04 S.Time : 14:30:03.20 E.Date : 2006-03-04 E.Time : 14:51:06.03 Host : MEDIA_HOLE
Trascription Verify		no						
Web Streaming		yes						
Streaming Profile		rai.prx						
Semantic Analysis		yes						
VCA Analysis		no						
Web Publication		yes						
Web Indexing		yes						
<input checked="" type="checkbox"/>	RAI3	TGR Piemonte	2006-3-04 13:59:58	2006-3-04 14:18:39	ANTS_1141661580.53	100.0%	complete	<input type="checkbox"/>
<input checked="" type="checkbox"/>	CANALE5	TG5	2006-3-04 12:58:36	2006-3-04 13:28:42	ANTS_1141661242.23	100.0%	complete	<input type="checkbox"/>
<input checked="" type="checkbox"/>	RAI2	TG2	2006-3-04 12:58:34	2006-3-04 13:25:17	ANTS_1141661308.27	Unknown	fallout	<input type="checkbox"/>
<input checked="" type="checkbox"/>	RAI2	TG2	2006-3-03 20:29:55	2006-3-03 20:59:30	ANTS_1141654250.28	Unknown	fallout	<input type="checkbox"/>
<input checked="" type="checkbox"/>	RAI1	TG1	2006-3-03 19:59:26	2006-3-03 20:31:45	ANTS_1141654019.47	100.0%	active	<input type="checkbox"/>
<input checked="" type="checkbox"/>	RAI3	TGR Piemonte	2006-3-03 19:35:20	2006-3-03 19:54:07	ANTS_1141661238.75	100.0%	complete	<input type="checkbox"/>
<input checked="" type="checkbox"/>	RAI3	TG3	2006-3-03 18:59:51	2006-3-03 19:31:16	ANTS_1141653919.46	100.0%	complete	<input type="checkbox"/>
<input checked="" type="checkbox"/>	RAI3	TG3	2006-3-03 14:24:23	2006-3-03 14:51:27	ANTS_1141653826.6	20.8%	complete	<input type="checkbox"/>
<input checked="" type="checkbox"/>	RAI3	TGR Piemonte	2006-3-03 14:00:57	2006-3-03 14:19:29	ANTS_1141653729.74	100.0%	complete	<input type="checkbox"/>
<input checked="" type="checkbox"/>	CANALE5	TG5	2006-3-03 12:57:32	2006-3-03 13:25:25	ANTS_1141653658.11	100.0%	complete	<input type="checkbox"/>
<input checked="" type="checkbox"/>	RAI2	TG2	2006-3-02 20:30:05	2006-3-02 21:00:00	ANTS_1141653640.36	Unknown	fallout	<input type="checkbox"/>
<input checked="" type="checkbox"/>	RAI3	TGR Piemonte	2006-3-02 19:33:48	2006-3-02 19:52:34	ANTS_1141653235.84	100.0%	complete	<input type="checkbox"/>
<input checked="" type="checkbox"/>	RAI3	TG3	2006-3-02 18:59:55	2006-3-02 19:29:58	ANTS_1141652160.24	100.0%	complete	<input type="checkbox"/>
<input checked="" type="checkbox"/>	RAI3	TG3	2006-3-02 14:23:28	2006-3-02 14:46:38	ANTS_1141312481.81	100.0%	complete	<input type="checkbox"/>
<input checked="" type="checkbox"/>	RAI3	TGR Piemonte	2006-3-02 14:00:04	2006-3-02 14:18:54	ANTS_1141653254.93	100.0%	complete	<input type="checkbox"/>
<input checked="" type="checkbox"/>	RAI1	TG1	2006-3-02 13:30:17	2006-3-02 13:57:21	ANTS_1141653130.51	100.0%	active	<input type="checkbox"/>
<input checked="" type="checkbox"/>	RAI2	TG2	2006-3-02 12:58:24	2006-3-02 13:30:35	ANTS_1141652810.97	67.2%	complete	<input type="checkbox"/>
<input checked="" type="checkbox"/>	CANALE5	TG5	2006-3-02 12:57:25	2006-3-02 13:30:43	ANTS_1141652716.39	97.8%	active	<input type="checkbox"/>
<input checked="" type="checkbox"/>	RAI3	TGR Piemonte	2006-3-01 19:33:44	2006-3-01 19:52:28	ANTS_1141652566.03	100.0%	complete	<input type="checkbox"/>
<input checked="" type="checkbox"/>	RAI3	TG3	2006-3-01 19:00:10	2006-3-01 19:29:41	ANTS_1141652235.33	100.0%	complete	<input type="checkbox"/>

Copyright © 2005 by RAI Crit & RAI Teche, Inc. All rights reserved

Centro Ricerche e Innovazione Tecnologica



Browse interface

ANTS Web Navigation - Microsoft Internet Explorer

Rai

Today is : 2006-09-20 12:38:38

Product Version 2.0b

Info Transcription Semantic VideoContent Linked News

Search FIND

19:00:11 il tentativo di colpo e' stata in thailandia militare casa del governo carri armati nelle strade il premier dichiara lo stato d'emergenza il leader iraniano smorza le polemiche contro il papa il sangue versato dice pontefice insieme di speranza per il rispetto reciproco sara' prodi e riferire in aula sul caso telecom alla camera l' opposizione meglio tardi che mai bertinotti rispettato il ruolo delle camere bum bum a palazzo chigi incontro con le parti sociali sulla finanziaria due mila sette riduzione nebulose scale

19:00:42 indici miliardi per lo sviluppo per luce e gas le famiglie spenderanno cento sessanta euro in piu' secondo federconsumatori sono gli aumenti piu' forte degli ultimi dieci anni forse un suicidio a gaza ha provocato il crollo della palazzina di milano sono morte quattro persone sospette su una delle vittime dopo le dimissioni di guido rossi e giovedi' il nuovo commissario della federcalcio il presidente del cono ironizza diremo nulla

19:01:11 arriva il nuovo album di zucchero fly per volare con usa e cubani dice il cantante nel oda anche bravi giovanotti cosa

19:01:23 la sera tutti dal tg3 e allora avete sentito dal mondo islamico toni piu' pacati dopo la protesta che era stata innescata dalle parole pronunciate da benedetto sedicesimo l' universita' tedesca diretti sbornia il presidente iraniano a forme di ineggiandi dice rispetto

19:01:41 me tutte le persone di pace d' altra parte lui stesso ha spiegato che le sue parole sono state male interpretate il papa deplora ogni forma di violenza ricorda suo rione lei dice che dal suo sangue versato possa nascere il si' di dida sentiamo vita ma ricca

19:02:00 caracas venezuela all' aeroporto un attimo prima di partire per majorca verso l' assemblea generale delle nazioni unite il presidente iraniano ma nuda ma di richard va al microfono gnocchi noi rispettiamo il patto e tutti devono essere interessati alla baci forma uguale sostanza soprattutto se arriva della torre

TIME 19:00:01:28

RAI3 - TG3 - 2006-09-19

- Politica, Partiti, Istituzioni e Sindacati
- Vita e Cultura Religiosa
- Esterni
- Politica, Partiti, Istituzioni e Sindacati
- Politica, Partiti, Istituzioni e Sindacati
- Politica, Partiti, Istituzioni e Sindacati
- Trasporti
- Consumi e Servizi
- Cronaca
- Sport
- Giustizia, Criminalità e Sicurezza
- Cronaca
- Other Classification
- Musica e Spettacolo
- Other Classification

TIMELINE

18:59:55:71

19:00:11:83

19:00:14:93

19:00:16:50

19:00:18:37

19:00:19:77

19:00:20:26

19:00:22:67

19:00:24:35

19:00:28:27

19:00:28:78

19:00:30:66

19:00:33:02

19:00:36:77

19:00:37:25

19:00:40:25

19:00:42:43

19:00:45:27

19:00:45:74

19:00:47:15

19:00:48:51

19:00:51:65

19:00:53:69

19:00:54:21

ANTS Copyright © 2005 by RAI Crit & RAI Teche, Inc. All rights reserved

Start Posta in arrivo - Microsoft... Multimedia Catalogue - A... Immagini ANTS_115.ppt [Sola lettura] ANTS Copyright © 2006... ANTS Web Navigation... Documenti 12.38



ANTS exploitation

- ✓ Predocumentation service
 - ☞ Regional productions
 - ☞ National newscasts
- ✓ Logging of all national Tv and Radio RAI channels
- ✓ Indexing and selection of content for new media publication
- ✓ Channel monitoring
 - ☞ Local stations monitoring service provided to Piedmont and Veneto local Communication Authorities
- ✓ EU projects
 - ☞ PrestoSpace <http://www.prestospace.eu>
- ✓ Commercialization in partnership with system integrators



PrestoSpace

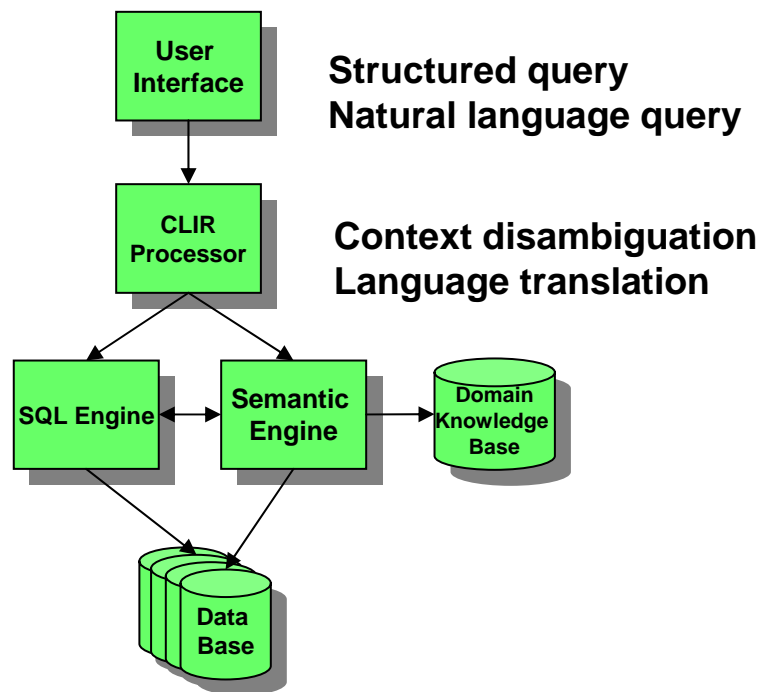
- ✓ IST 6th Framework Integrated Project
- ✓ It's main goal is the study and design of **preservation factories** for the digitisation, restoration and documentation of audiovisual content
- ✓ Main partners are: INA (prime contractor), RAI, BBC, ORF, B&G, Joanneum Research
- ✓ The project will end in January 2008



PrestoSpace MAD

RAI is responsible for the documentation Area,
called MAD (Metadata, Access and Delivery)

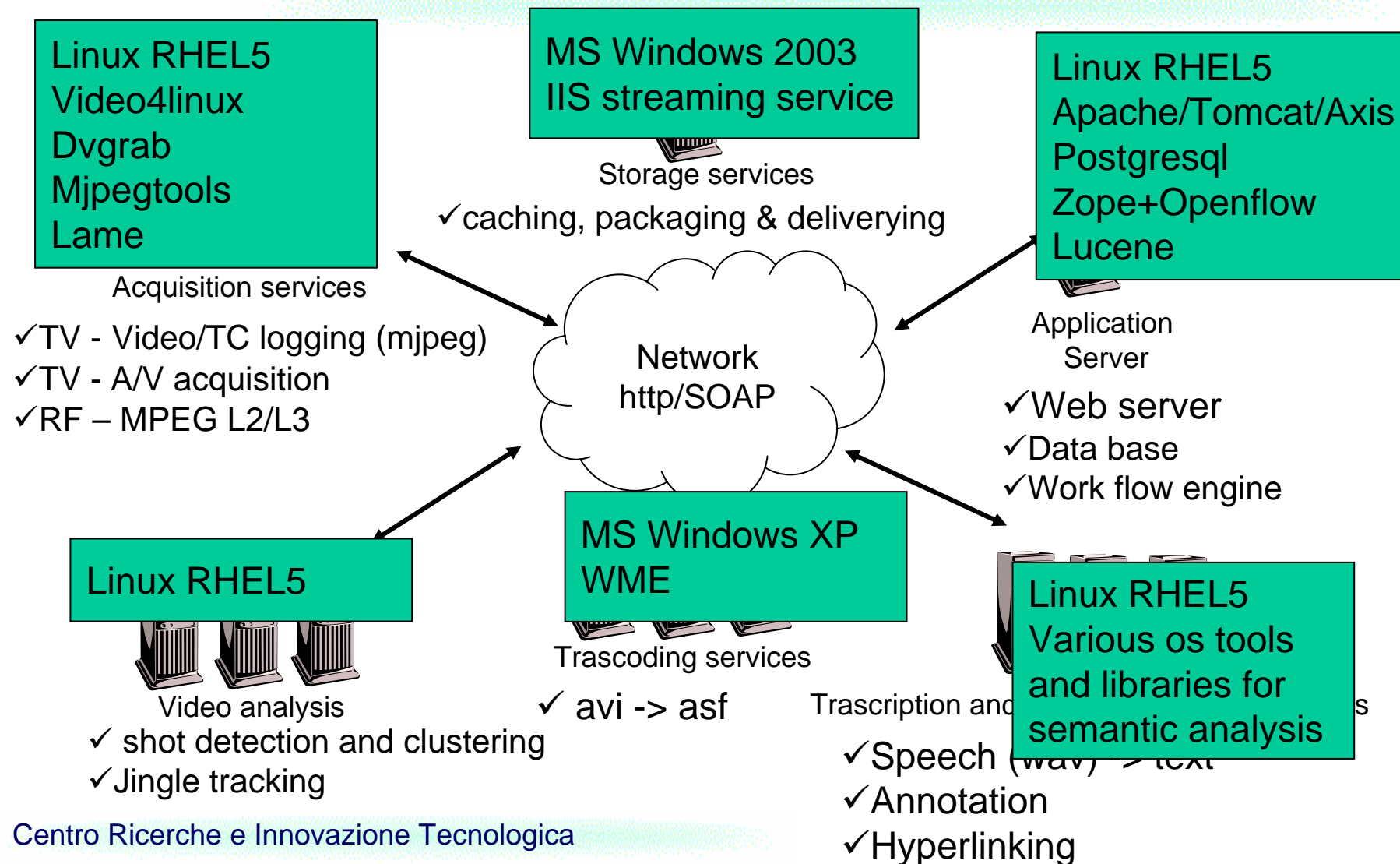
RETRIEVAL OF CONTENT



- ✓ The documentation process is based on ANTS technologies
- ✓ The publishing platform includes a semantic search engine using a domain specific onthology
- ✓ The users can express their queries using their natural language
- ✓ A query preprocessor (CLIR) analyses the query, exctracts the most relevant concepts and applies language translation if necessary



ANTS basic software





linux

- ✓ We use Fedora Core for prototyping and Red Hat Enterprise Linux for production
- ✓ Maximum configuration flexibility
- ✓ Good overall stability but
 - ☞ Sometime problems with newer hardware
 - Requires kernel update/options hacking
- ✓ Some peripheral drivers not included in standard kernel
 - ☞ Need to be manually compiled
- ✓ Difficulty in handling proprietary closed source drivers
- ✓ Advanced system knowledge required



Video capture

- ✓ High quality video capture
 - ☞ Few pro cards supported (more are coming)
 - ☞ DV capture via firewire ok
- ✓ Low resolution video
 - ☞ Supported via the video4linux project
 - ☞ Many TV capture cards available at very low cost (bt878 chip based)
 - ☞ Lots of variants and revisions cause random problems during deployment
- ✓ Very few card drivers are provided by their manufacturers
- ✓ We are switching to Windows and Direct Show to increase the choice of cards and formats



Audio capture

- ✓ Alsa and Jack projects provide a stable and performing environment for professional audio
- ✓ Modified kernel can be necessary for low latency and uninterrupted performance
- ✓ Limited choice of pro cards
- ✓ Better firewire support would be welcome



Middleware and db

- ✓ Wide variety of high quality middleware tools including connectivity, programming languages and libraries
- ✓ Projects like Apache, Tomcat, Lucene and postgresSQL allow for developement of high performing professional applications
- ✓ Free availability and public knowledge base are essential for fast evaluation, prototyping and bug fixing



Scientific libraries and tools

- ✓ Several universities and research centres are providing in open source advanced tools and libraries for image, sound and text processing
- ✓ We are using some for automatic classification based on content features
 - 👉 CPAN Categorizer, LIBSVM
- ✓ We are also experimenting with open source speech to text engines
 - 👉 Sphinx, ISIP

Conclusions

- ✓ Linux and many open source projects can be successfully used in professional quality audiovideo applications
- ✓ Free availability, public knowledge base, source code hacking possibilities
- ✓ Advanced technical knowledge often required
- ✓ Better support for specific audio and video hardware is needed
- ✓ The policy for using non free codecs (e.g. mp3) in commercial environments should be clarified