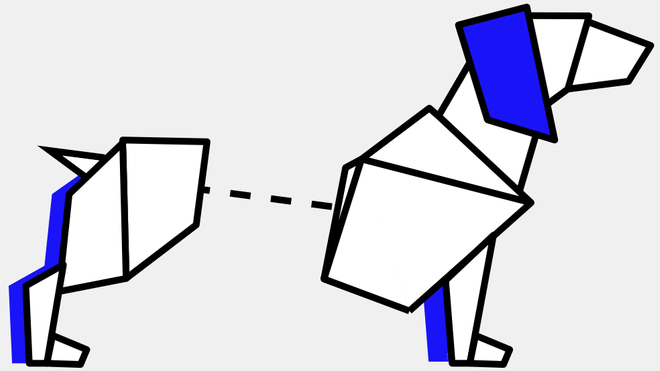


LINKED DATA IN FILM ARCHIVES

ACE Workshop, Bologna, 27 June 2017

TABLE OF CONTENTS

Chapter : Cinémathèque royale de Belgique, **a brief history**

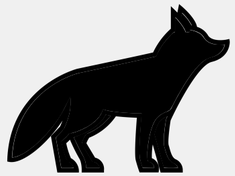
Chapter : Linked Data, **a practical introduction**

Chapter : Cinémathèque royale de Belgique, **a brief migration**

Chapter : Cinémathèque royale de Belgique, **a brief demo**

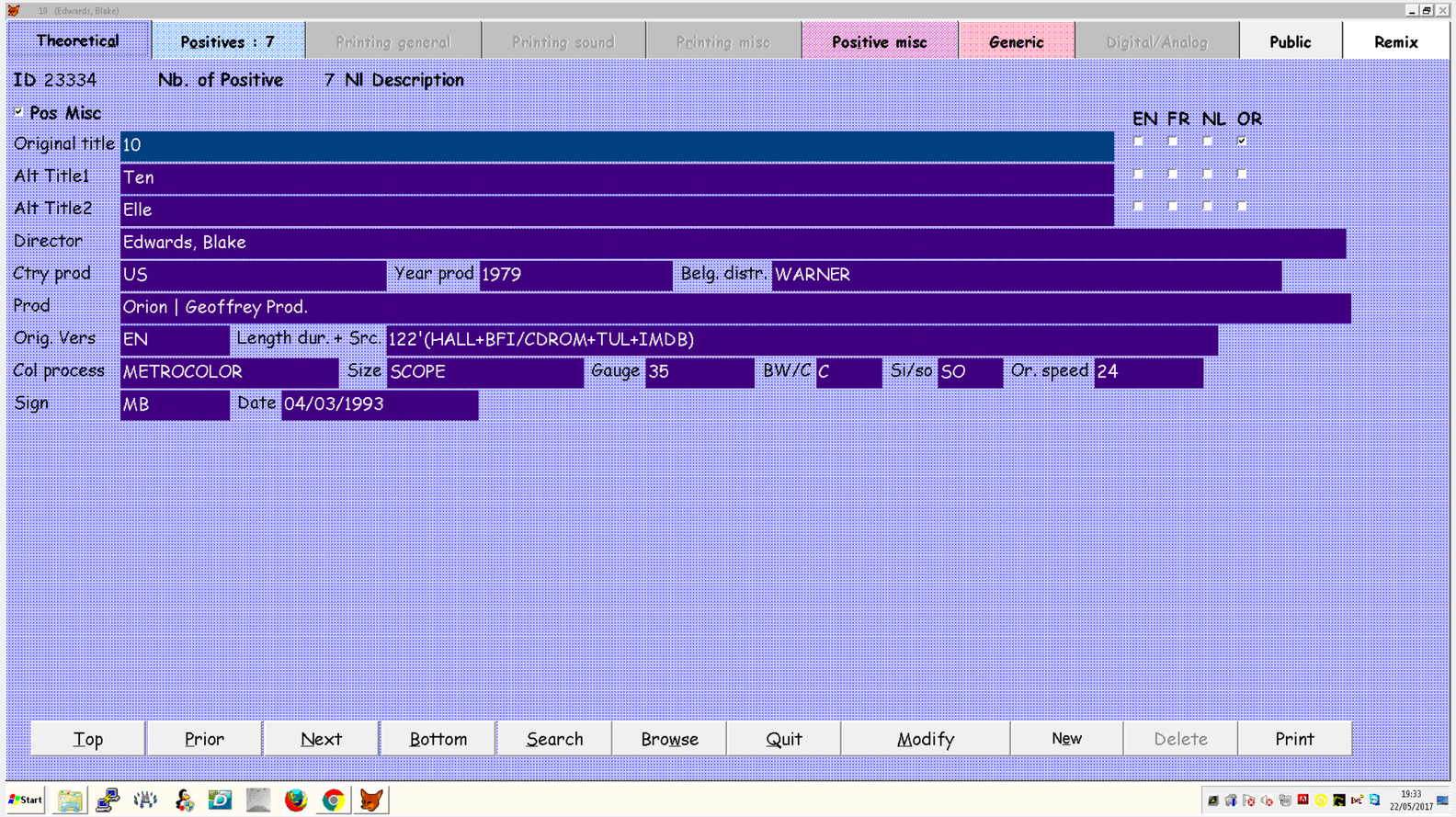
Chapter : Linked Data, **the next frontier**

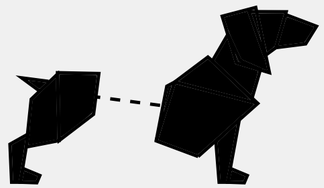
Chapter : Cinémathèque royale de Belgique, **the ontology**



Chapter one
Cinémathèque royale de Belgique
a brief history

born in the 90's
support ended
slow
errors, crashes
no standard at all


10 (Edwards, Blake)
Theoretical
Positives : 7
Printing general
Printing sound
Printing misc
Positive misc
Generic
Digital/Analog
Public
Remix
ID 23334
Nb. of Positive
7 NI Description
Pos Misc
EN FR NL OR
Original title 10
Alt Title1 Ten
Alt Title2 Elle
Director Edwards, Blake
Ctry prod US
Year prod 1979
Belg. distr. WARNER
Prod Orion | Geoffrey Prod.
Orig. Vers EN
Length dur. + Src. 122'(HALL+BFI/CDROM+TUL+IMDB)
Col process METROCOLOR
Size SCOPE
Gauge 35
BW/C C
Si/so SO
Or. speed 24
Sign MB
Date 04/03/1993
Top
Prior
Next
Bottom
Search
Browse
Quit
Modify
New
Delete
Print
Start
19:33
22/05/2017



Chapter two
Linked Data
a practical introduction

Linked Data is easy: only three words you should know!

Triple

Like in
one, two, triple

URI

The address-thing
in your internet
browser

Ontology

Google says:
a set of concepts
and categories

Linked Data knowledge is modeled in "**triples**"

1. Subject	**2. Predicate**	**3. Object**
D.O.A.	is a	film
D.O.A.	published in	1949
D.O.A.	has director	Rupolph Mate
D.O.A.	written by	Russell Rouse
D.O.A.	genre	film noir
D.O.A.	still picture	

• • • http://www.cinematek.be/resource/Film/877920329832560392

Triples are expressed in **URI**s, short for "Uniform Resource Identifiers"

Subject	**Predicate**	**Object**
D.O.A. http://www.cinematek.be/resource/Film/877920329832560392	is a http://www.w3.org/1999/02/22-rdf-syntax-ns#type	film http://www.cinematek.be/ontology/Film

↑

Note: there's no "ID", the entire URI is the ID

"OK, so the URI is the unique identifier of the film,
but what if two institutions have the same film,
how can it be unique?"

CINEMATEK

http://www.cinematek.be/resource/Film/D.O.A.

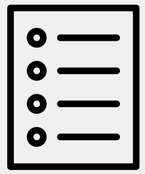
http://www.cinetecadibologna.it/resource/Film/D.O.A.

It's linked data, so we just "link" them together

no need for a third unique identifier party like ISAN or EIDR!

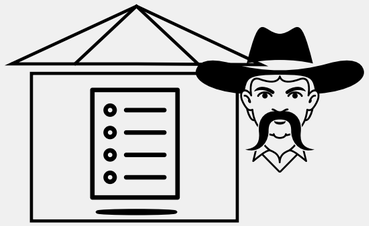
Subject	Predicate	Object
http://www.cinematek.be/resource/Film/D.O.A.	owl:sameAs	http://www.cinetecadibologna.it/resource/Film/D.O.A.

Triples are organized in "**ontologies**"



a collection

with a home

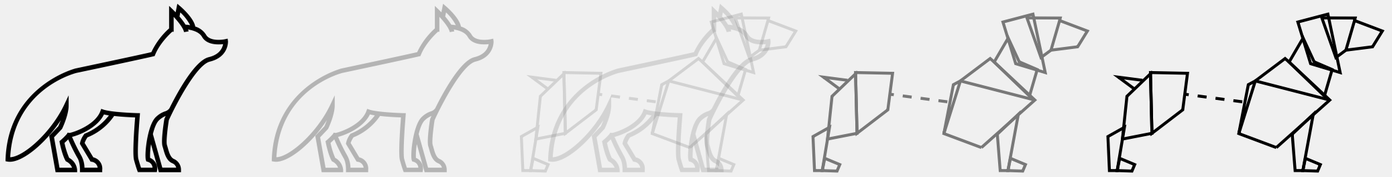


and a boss
(with rules)

Chapter three
Cinémathèque royale de Belgique
a brief migration

CINEMATEK

the future of FoxPro: Linked (Open) Data



Phase

From a two- to three-layered datamodel

- Theatrical release information
  les "fiches théoriques"
- Storage information
  negatives, positives, optical, tape, digital, ...
- -

→

- CWS level 1: Work
- CWS level 2: Manifestation
- CWS level 3: Item

Phase

Define missing fields

What about fields missing from EN 15907?

Length of a silent movie without intertitles

Good/bad quality of copy

Extra flags
- for museum
- for EFG

FORWARD copyright info

Phase

Data cleanup

1	TCHNICOLOR
1	GTC COLOR
1	BW+FILTERS
1	DUFAY+GASPAR
1	CHEMICOLOR
1	ANSCOCHROME
1	CFI|DELUXE
1	ESTMANCOLOR
1	KELLER-DORIAN
1	NE
1	DUART COLOR
1	FERRANIA
1	ECLAIRCOLOR
1	DELUX
1	KINECOLOR
1	DUNNINGCOLOR
1	DE LUXE|DUART
1	RANKCOLOR !
1	PATHE COLOR]
2	UNITED COLOR
2	CINECITTA
2	!
2	-

31	PATHECOLOR
33	SOVCOLOR
35	DU ART
37	EKTACHROME
38	KODACHROME
39	CINECOLOR
41	TRUCOLOR
46	FOTOKEM
55	RANK COLOR
60	MOVIELAB
63	DUART
72	FUJICOLOR
92	AGFACOLOR
96	WARNERCOLOR
118	GEVACOLOR
238	CFI COLOR
357	EASTMANCOLOR
403	METROCOLOR
517	DELUXE
635	EASTMANCOLO
769	DE LUXE
3399	TECHNICOLOR
70323	

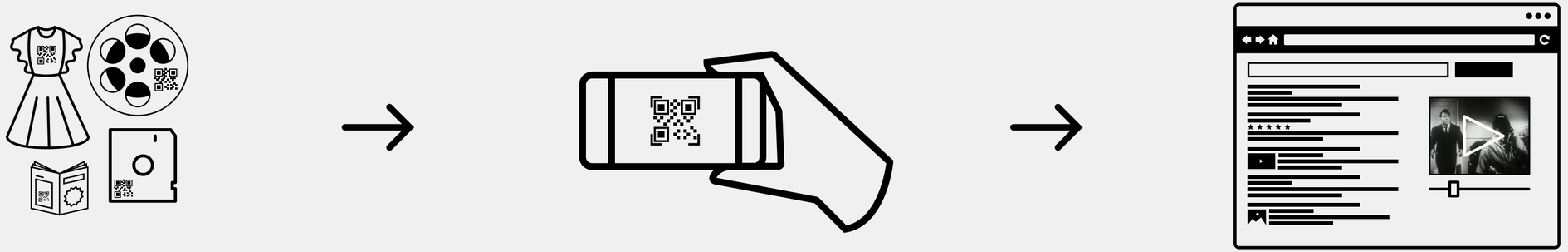
Extract of the color Systems @CINEMATEK

Phase

Integrate into workflows

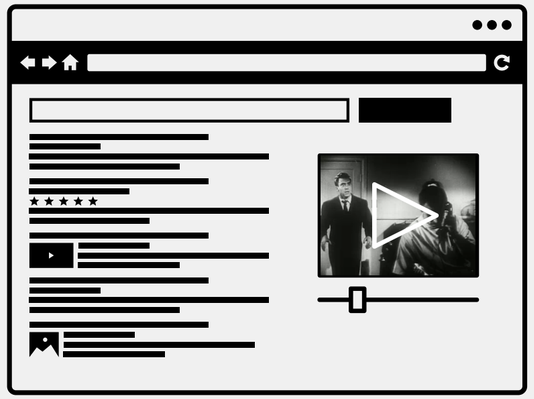
http://www.cinematek.be/resource/Film/D.O.A.

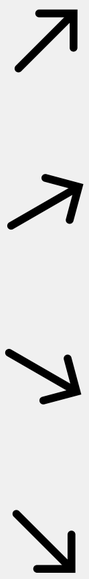
In the end, the ID of the film is just a web address,
so you can encode and open it on any device you want



Phase

Link to other ontologies





Non-film collections

Geonames

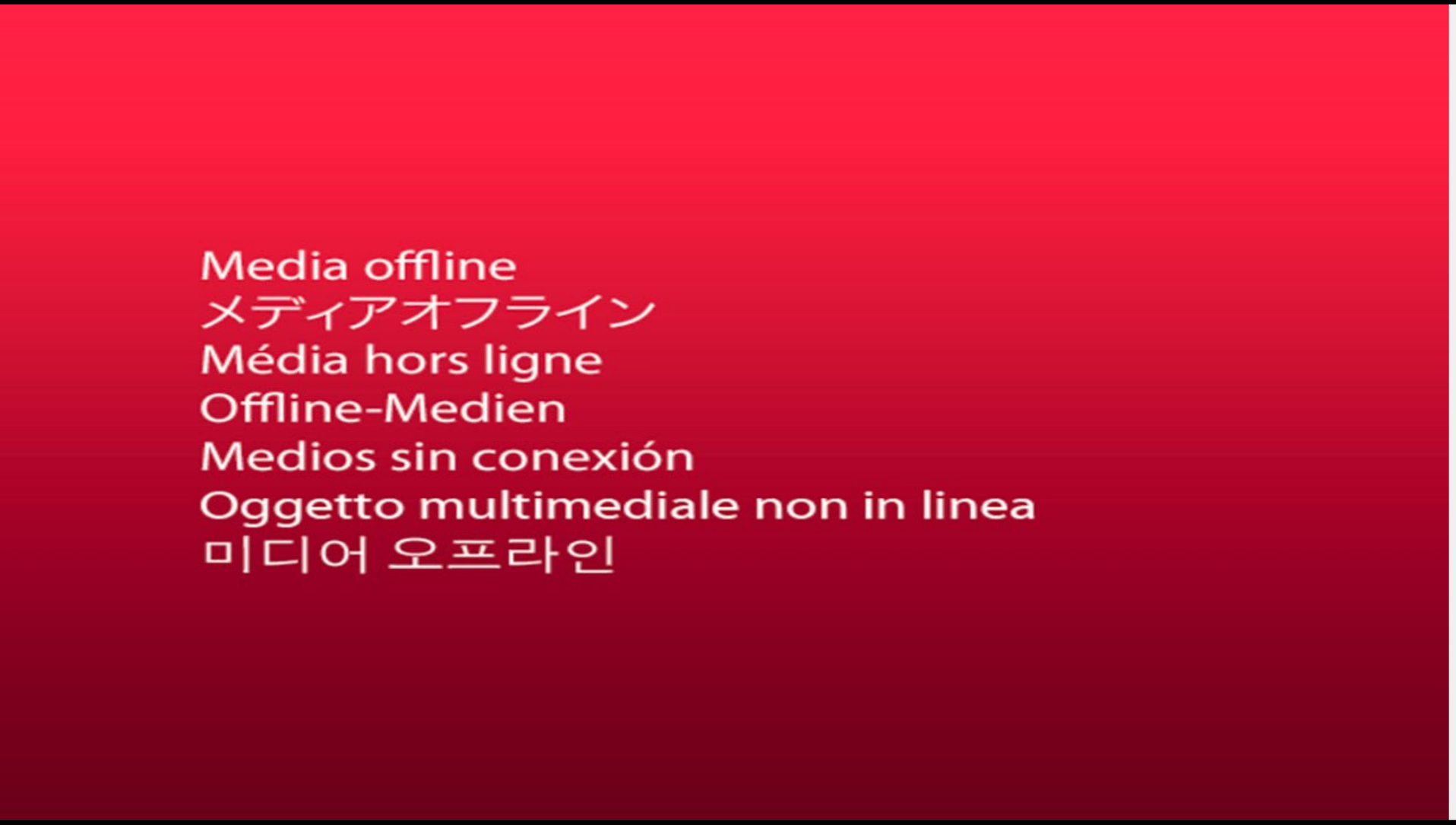
Wikipedia

...

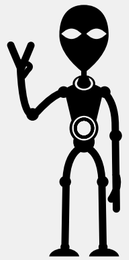
VIAF/Getty/LoC/...

Chapter four

Cinémathèque royale de Belgique
a brief demo



Note
Contact Bram if you want to get your hands on this video
(see last slide)



Chapter five
Linked Data
the next frontier

What happens when a **computer** looks at a video file?

General

Count :	324
Count of stream of this kind :	1
Kind of stream :	General
Kind of stream :	General
Stream identifier :	0
Inform :	MPEG-4 (Base Media / Version 2): 22.7 MiB, 2mn 14s
Count of video streams :	1
Count of audio streams :	1
Video_Format_List :	AVC
Video_Format_WithHint_List :	AVC
Codecs Video :	AVC
Audio_Format_List :	AAC
Audio_Format_WithHint_List :	AAC
Audio codecs :	AAC LC
Complete name :	/home/bram/Projects/cinematek/bologna/media/videos/Elvea Pomo e Legumi - cli
Folder name :	/home/bram/Projects/cinematek/bologna/media/videos
File name :	Elvea Pomo e Legumi - clip 1 - NL-BaU1rPhgzUw
File extension :	mp4
Format :	MPEG-4
Format :	MPEG-4
Format/Extensions usually used :	mp4 m4v m4a m4b m4p 3gpp 3gp 3gpp2 3g2 k3g jpm jpx mqv ismv isma f4v
Commercial name :	MPEG-4
Format profile :	Base Media / Version 2
Internet media type :	video/mp4
Codec ID :	mp42
Codec ID :	mp42 (isom/mp42)
Codec ID/Url :	http://www.apple.com/quicktime/download/standalone.html
CodecID_Compatible :	isom/mp42
Codec :	MPEG-4
Codec :	MPEG-4
Codec/Extensions usually used :	mp4 m4v m4a m4b m4p 3gpp 3gp 3gpp2 3g2 k3g jpm jpx mqv ismv isma f4v
File size :	23788637
File size :	22.7 MiB
File size :	23 MiB
File size :	23 MiB
File size :	22.7 MiB
File size :	22.69 MiB
Duration :	134258

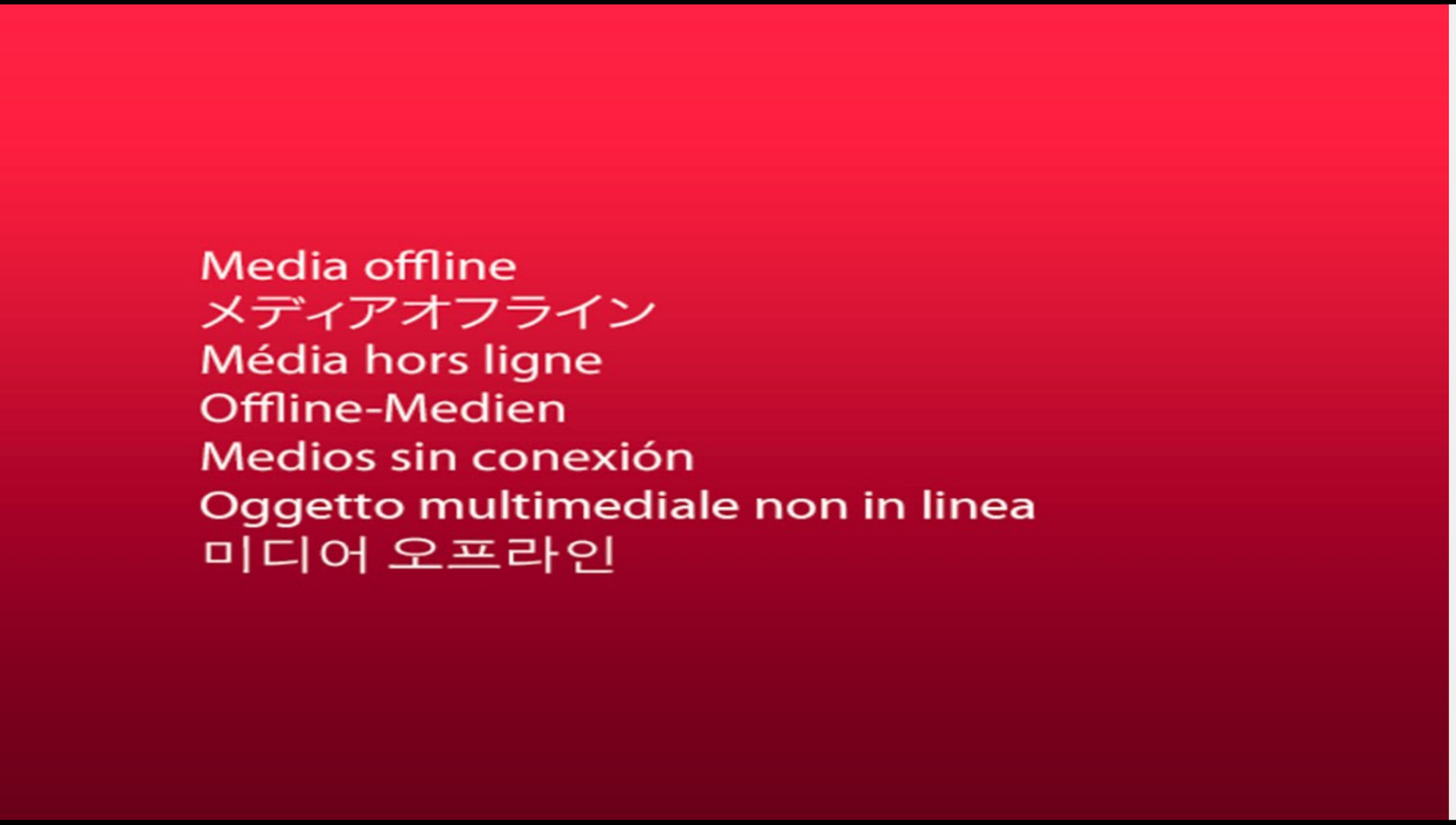
Video

Count :	332
Count of stream of this kind :	1
Kind of stream :	Video
Kind of stream :	Video
Stream identifier :	0
StreamOrder :	0
Inform :	1 289 Kbps, 1280*720 (16:9), at 25.000 fps, AVC (Main@L3.1) (CABAC / 3 Ref Frames)
ID :	1
ID :	1
Format :	AVC
Format/Info :	Advanced Video Codec
Format/Url :	http://developers.videolan.org/x264.html
Commercial name :	AVC
Format profile :	Main@L3.1
Format settings :	CABAC / 3 Ref Frames
Format settings, CABAC :	Yes
Format settings, CABAC :	Yes
Format settings, ReFrames :	3
Format settings, ReFrames :	3 frames
Internet media type :	video/H264
Codec ID :	avc1
Codec ID/Info :	Advanced Video Coding
Codec ID/Url :	http://www.apple.com/quicktime/download/standalone.html
Codec :	AVC
Codec :	AVC
Codec/Family :	AVC
Codec/Info :	Advanced Video Codec
Codec/Url :	http://developers.videolan.org/x264.html
Codec/CC :	avc1
Codec profile :	Main@L3.1
Codec settings :	CABAC / 3 Ref Frames
Codec settings, CABAC :	Yes
Codec_Settings_RefFrames :	3
Duration :	134240
Duration :	2mn 14s
Duration :	2mn 14s 240ms
Duration :	2mn 14s
Duration :	00:02:14.240

Audio

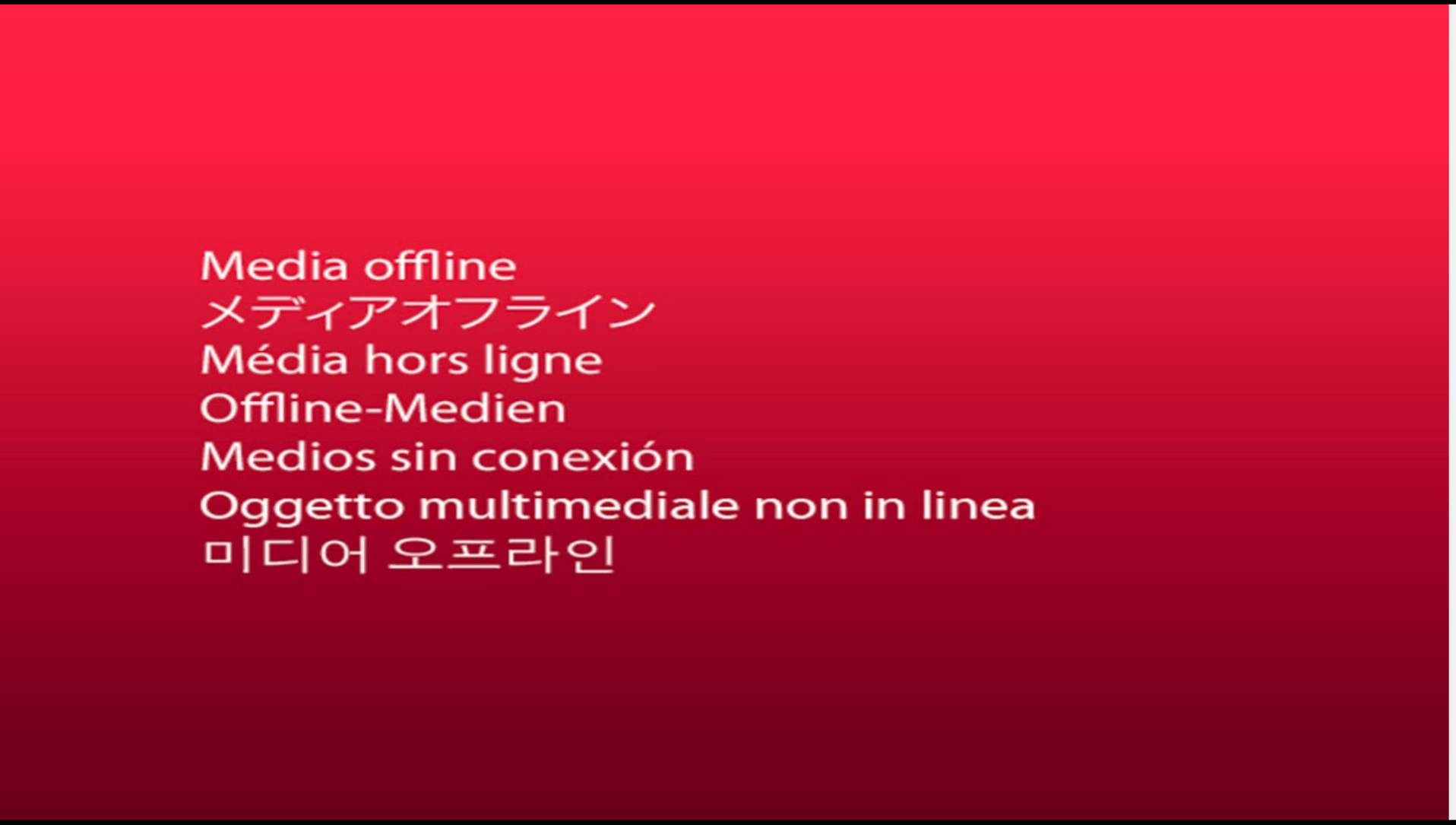
Count :	272
Count of stream of this kind :	1
Kind of stream :	Audio
Kind of stream :	Audio
Stream identifier :	0
StreamOrder :	1
Inform :	126 Kbps, 44.1 KHz, 2 channels, AAC (LC)
ID :	2
ID :	2
Format :	AAC
Format/Info :	Advanced Audio Codec
Commercial name :	AAC
Format profile :	LC
Codec ID :	40
Codec :	AAC LC
Codec :	AAC LC
Codec/Family :	AAC
Codec/CC :	40
Duration :	134258
Duration :	2mn 14s
Duration :	2mn 14s 258ms
Duration :	2mn 14s
Duration :	00:02:14.258
Duration :	00:02:14:20
Duration :	00:02:14.258 (00:02:14:20)
Bit rate mode :	VBR
Bit rate mode :	Variable
Bit rate :	125588
Bit rate :	126 Kbps
Channel(s) :	2
Channel(s) :	2 channels
Channel positions :	Front: L R
Channel positions :	2/0/0
ChannelLayout :	L R
Samples per frame :	1024
Sampling rate :	44100
Sampling rate :	44.1 KHz
Samples count :	5920778
Frame rate :	43.066
Frame rate :	43.066 fps (1024 spf)
Frame count :	5782

What happens when a **human** goes to the movies?



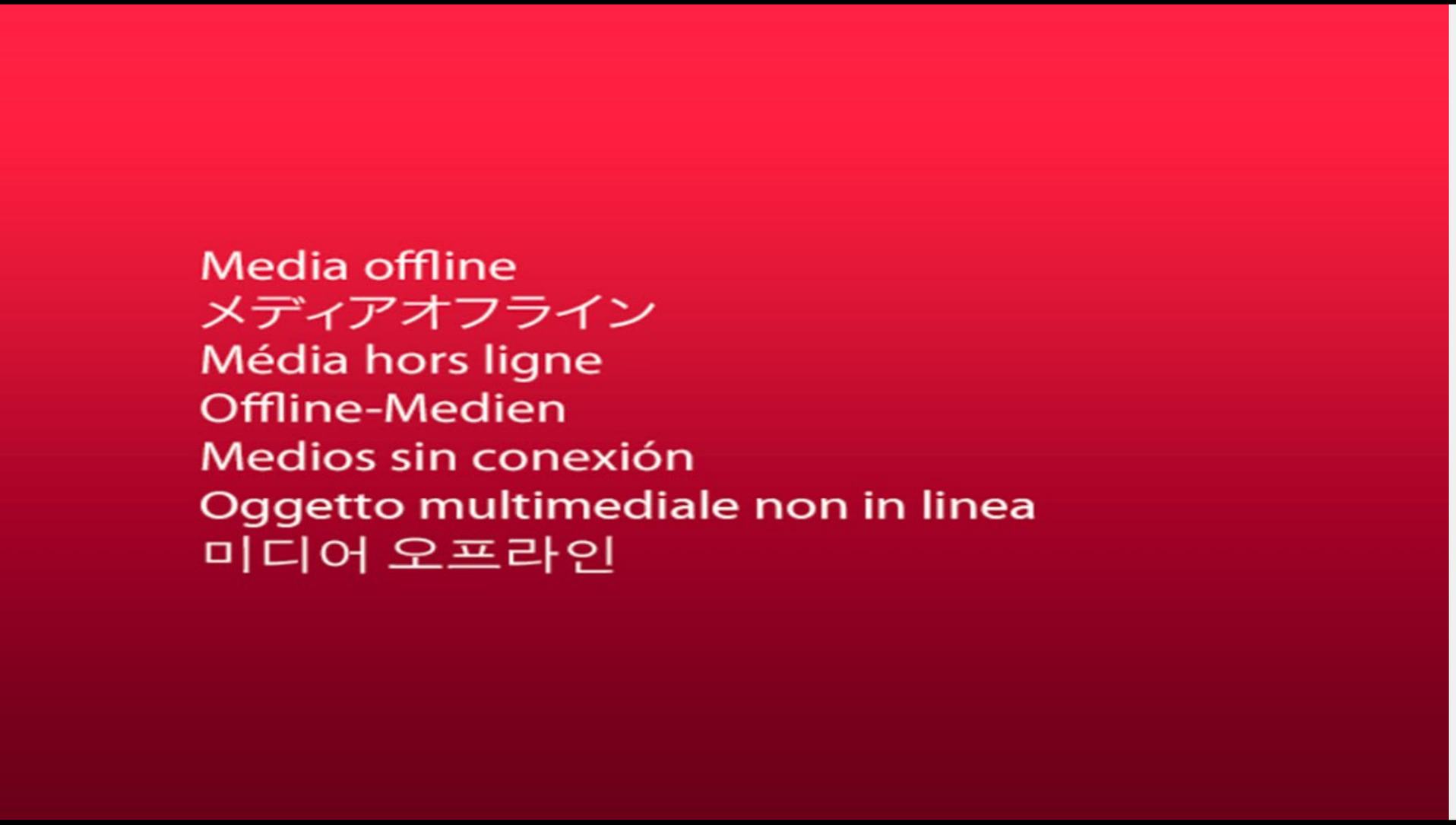
Note
Contact Bram if you want to get your hands on this video
(see last slide)

What happens when a **computer** goes to the movies?



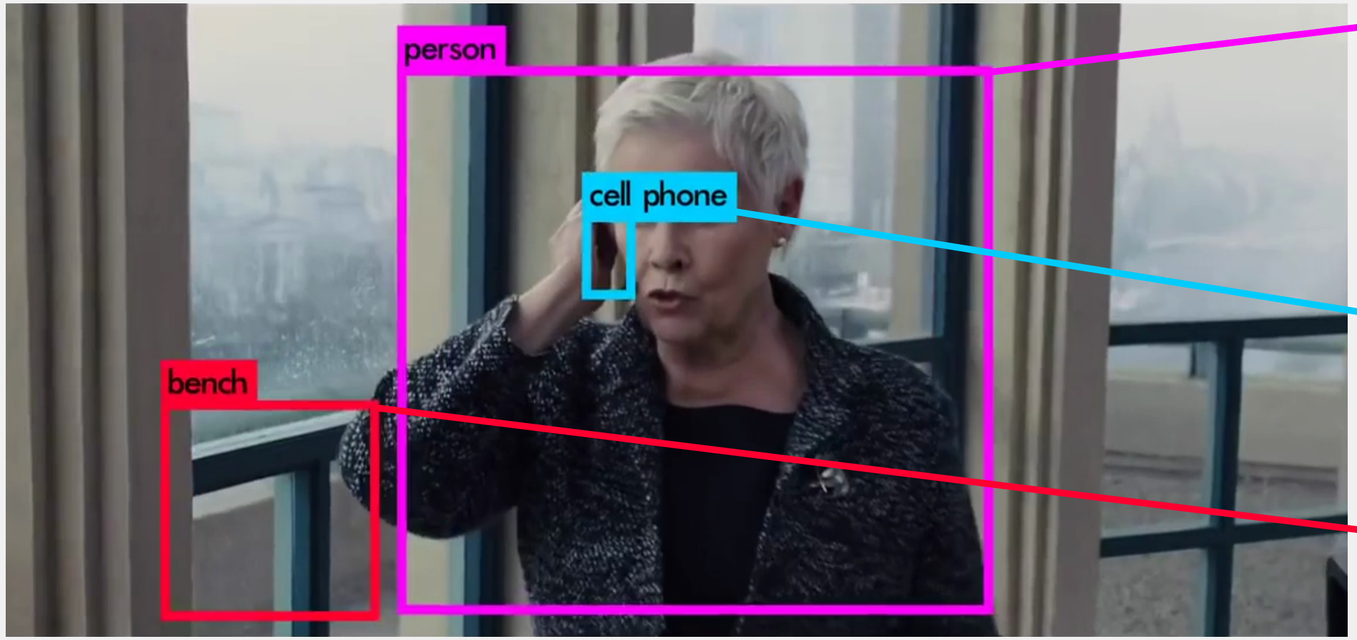
Note
Contact Bram if you want to get your hands on this video
(see last slide)

What happens when a **computer** decides to **read the words** in a movie?



Note
Contact Bram if you want to get your hands on this video
(see last slide)

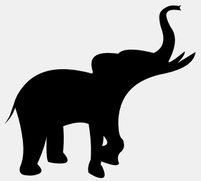
Automatic Analysis + Linked Data = POWAH!



https://en.wikipedia.org/wiki/Person
A person is a being, such as a human, that has certain capacities or attributes constituting personhood, which in turn is defined differently by different authors in different disciplines, and by different cultures in different times and places.

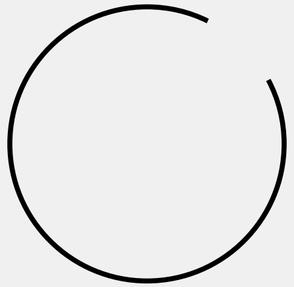
https://en.wikipedia.org/wiki/Mobile_phone
A mobile phone is a portable telephone that can make and receive calls over a radio frequency link while the user is moving within a telephone service area.

https://en.wikipedia.org/wiki/Bench_(furniture)
A bench is a long seat on which multiple people may sit at the same time.



Chapter six

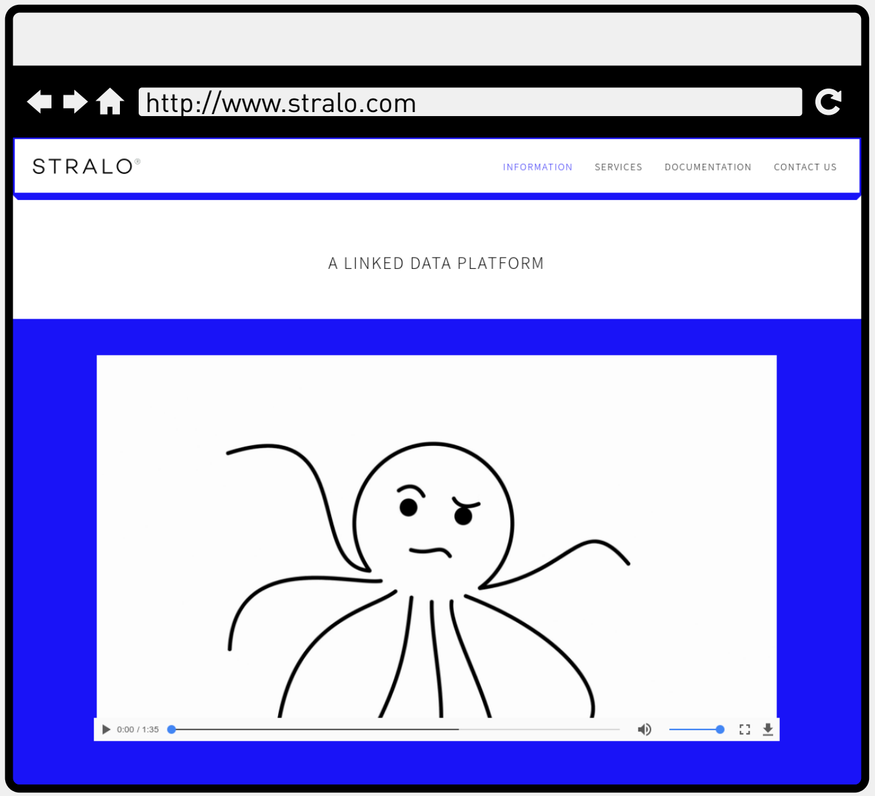
Cinémathèque royale de Belgique
the ontology



STRALO

A new Open Source, Linked Data project by

CINEMATEK | Republic of Reinvention



Website and documentation under construction

www.stralo.com

Aims of the project:

- open source software
- shared development
- free to use, commercial support if needed
- using local installation & infrastructure
- for all database types: film, non-film, media, ...
- attached to external Linked Data ontologies
- broadcast-competitive software quality

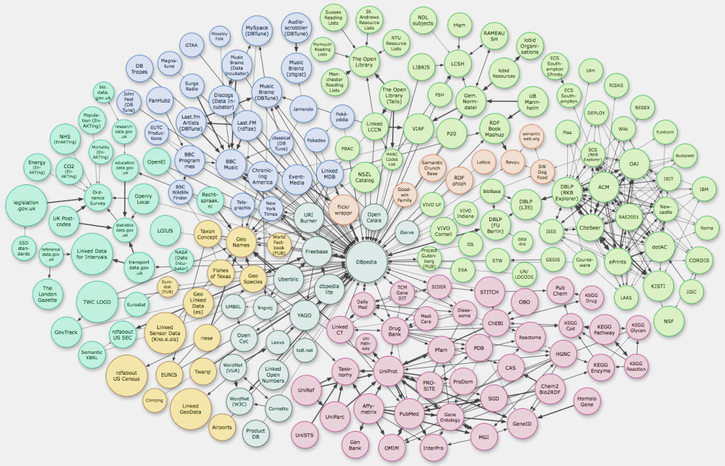
For the past few months, we've been translating this

The FIAF Moving Image Cataloguing Manual

Appendix K. List of Elements

WORK Section	WORK: Element Name	MANIFEST. Section	MANIFESTATION: Element name	ITEM Section	ITEM: Item Name
1.2.1	Work/Variant Description Type				
1.4.2	Variant Type	2.2.1	Manifestation Type		
1.3.1	Identifier	2.3.1	Identifier	3.1.1	Identifier
1.3.1.1	Identifier Type	2.3.1.1	Identifier Type	3.1.1.1	Identifier Type
1.3.2	Title	2.3.2	Title	3.1.2	Title
1.3.2.1	Title Type	2.3.2.1	Title Type	3.1.2.1	Title Type
1.4.1	Agent	2.4.1	Agent(s)	3.3.1	Agent(s)
1.4.1.1	Agent Activity	2.4.1.1	Agent Activity	3.3.1.1	Agent Activity
1.3.3	Country of Reference				
1.3.5	Language(s)	2.3.3	Language		
1.3.5.1	Language Term	2.3.3.1	Language Terms		
1.3.5.2	Usage Type	2.3.3.2	Usage Type		
1.3.4	Year/Date of Reference				
1.3.4.1	Date Type				
1.3.4.2	Date Precision				
1.4.3	Subject/Genre/Form Terms				
1.3.6	Content Description				
1.3.6.1	Content Description Type				
1.3.7	Notes	2.3.6	Notes	3.1.7	Notes
1.3.8	History				
1.3.8.1	Custodial History				
1.3.8.2	Censorship History				
1.3.8.3	Other Work/Variant History				
1.4	Relationships	2.4	Relationships	3.3	Relationships
1.4.2	Events	2.4.2	Events	3.3.2	Events
1.4.4	Other Relationships	2.4.3	Other Relationships	3.3.3	Other Relationships
		2.4.4	Items		
				3.1.3	Holding Institution
				3.1.4	Element Type
		2.3.4	Format	3.1.5	Item Physical+Digital Description
		2.3.4.1	Carrier Type	3.1.5.1	Carrier Type
		2.3.4.1.1	Carrier Type: General	3.1.5.1.1	Carrier Type: General
		2.3.4.1.2	Carrier Type: Specific	3.1.5.1.2	Carrier Type: Specific
		2.3.4.3	Sound Characteristics	3.1.5.3	Sound
		2.3.4.3.1	Sound systems	3.1.5.3	Sound System
		2.3.4.3.2	Sound Channel Configuration	3.1.5.4	Sound Channel Configuration
		2.3.4.4	Colour Characteristics	3.1.5.6	Colour
				3.1.5.7	Unit Number
		2.3.5	Extent		
		2.3.5.1	Logical Extent		
		2.3.5.2	Physical Extent	3.1.5.8	Extent
		2.3.4.2	Projection Characteristic	3.1.5.9	Projection Characteristics
				3.1.5.10	Broadcast Standard
		2.3.5.3	Duration	3.1.5.11	Duration
		2.3.5.3.4	Duration Precision	3.1.5.11.1	Duration Precision
		2.3.5.3.5	Duration Type		
				3.1.5.12	Frame Rate
				3.1.5.13	Base
				3.1.5.14	Stock
				3.1.5.15	Stock batch
				3.1.5.16	Video Codec
				3.1.5.17	Audio Codec
				3.1.5.18	Resolution
				3.1.5.19	Line Standard
				3.1.5.20	Bit Depth
				3.1.5.2	Item Status
				3.1.6.1	Item Condition
				3.1.6.2	Item Location
				3.1.5.21	Source Device
				3.1.5.22	Source Software
				3.1.5.23	Transfer Speed

into this



The CRB ontology

Three goals:

Create a CWS (EN 15907) -compatible ontology at Cinematek

During data cleanup, construct datasets of linked terms

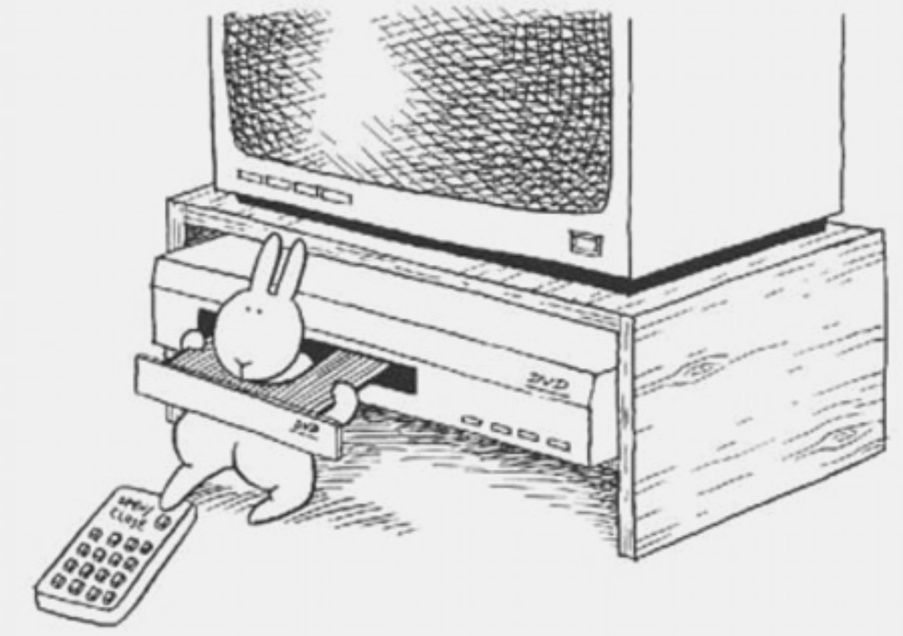
Open-source the above two

It's far from done, but we would like to share version 0.1 with you:

https://drive.google.com/open?id=1ctktopoJjQKzvlS0GprYr6zD5eWT-A7ItAawYSd-RSU

Note: datasets will follow shortly, please contact us if you're interested already.

The end

Bruno Mestdagh
bruno.mestdagh@cinematek.be

Bram Biesbrouck
bram.biesbrouck@reinvention.be