<Apache Tika/>

Content analysis for ECM with Apache Tika

Paolo Mottadelli - source**sense**

Leading the Wave of Open Source

paolo@apache.org

Apache	Sourcesense	Decathlon	Bobsleigh
Developing POI	Leading presales	Italian Champion	Olympic Team
Presenting Tika			
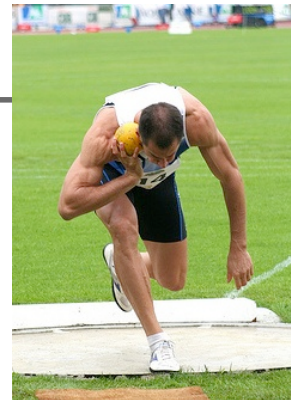
ON BOARD!

Paolo **Mottadelli**

Agenda

Some problems, one (possible) solution

Apache Tika: a presentation

ECM systems: some challenges

Tika for Alfresco: a use case

Supporting Open XML: a real life demo

Paolo **Mottadelli**

sourcesense

Main challenge

Lucene index

Agenda

Some problems, one (possible) solution

Apache Tika: a presentation

ECM systems: some challenges

Tika for Alfresco: a use case

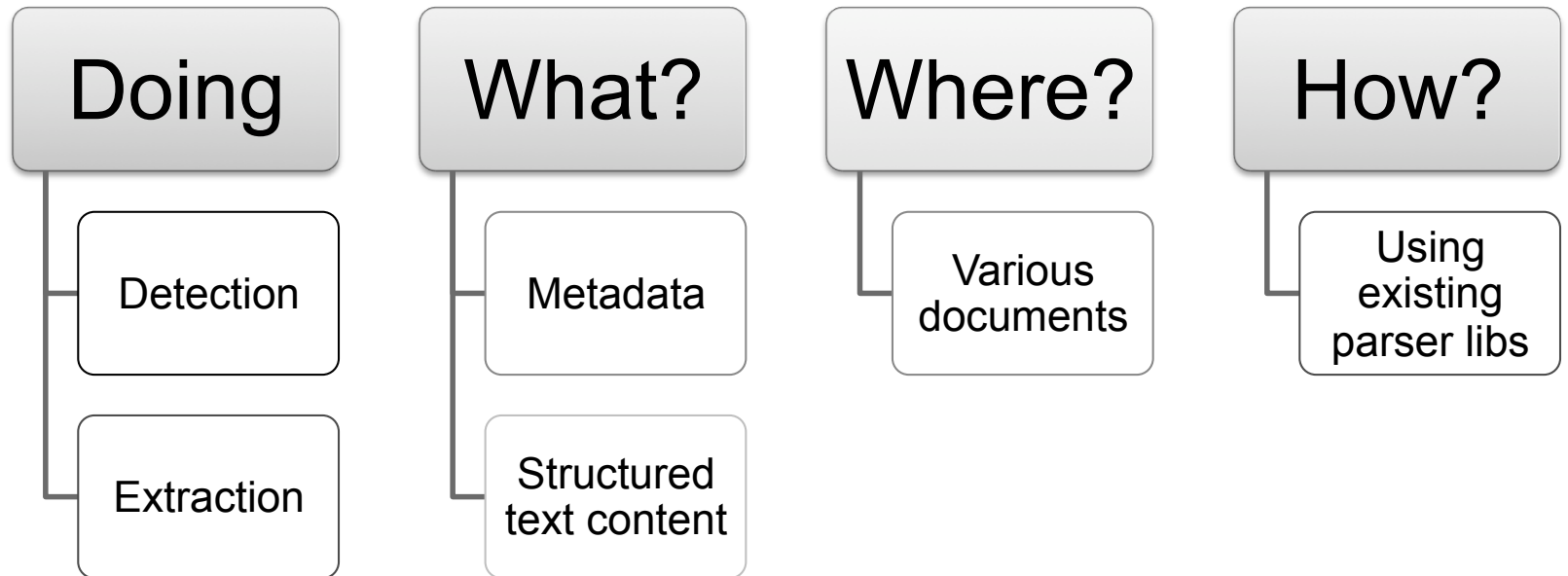Supporting Open XML: a real life demo

Paolo **Mottadelli**

sourcesense

Leading the Wave of Open Source
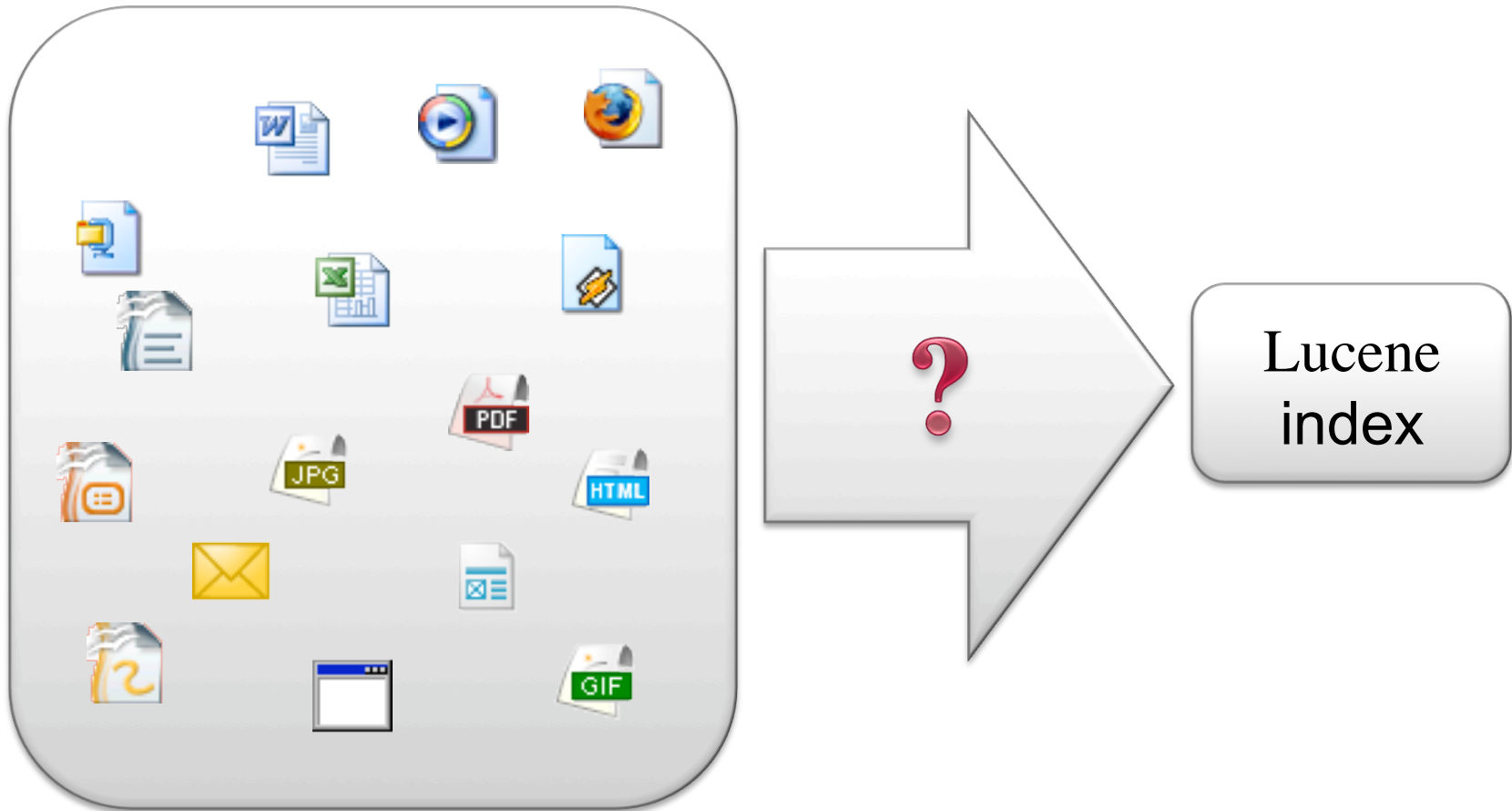
What is Tika?

Another Indian Lucene project? No.

ApacheCon

Leading the Wave of Open Source

Paolo **Mottadelli**   source**sense**

What is Tika?

It is a **Toolkit**

Doing	What?	Where?	How?
Detection	Metadata	Various documents	Using existing parser libs
Extraction	Structured text content		

To be continued…

ApacheCon

p.mottadelli@sourcesense.com

Leading the Wave of Open Source

Paolo **Mottadelli**

source**sense**