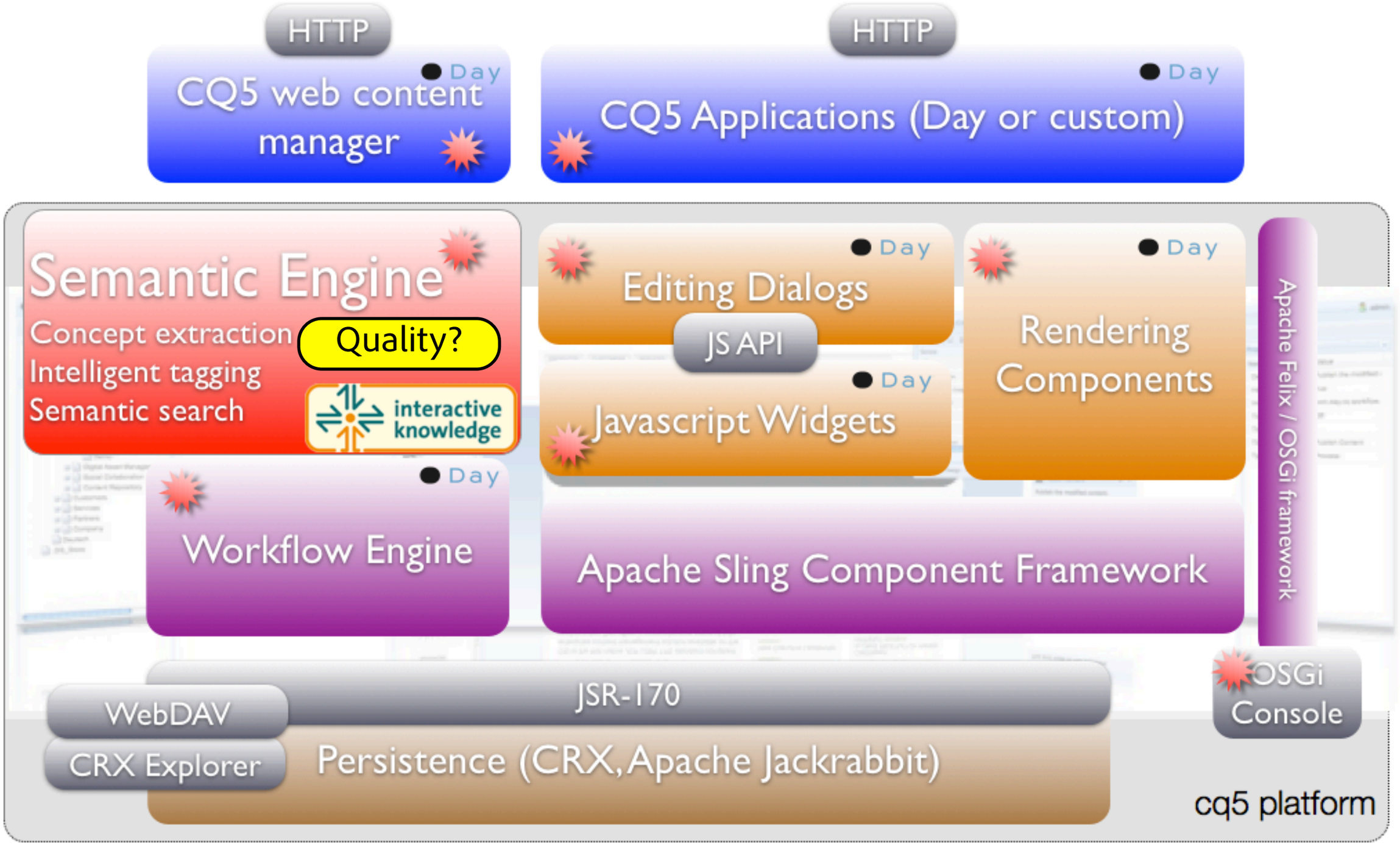
Apache Stanbol Enhancer Benchmarks

Bertrand Delacretaz, Senior R&D Developer, Adobe Systems
IKS workshop + review, Saarbrücken, April 2011 - slides version 2011-04-08

© 2011 Adobe Systems Incorporated. All Rights Reserved. Adobe Confidential.

New year, same vision



OSGi "plugins

Day

© 2011 Adobe Systems Incorporated. All Rights Reserved.

Benchmarks: evaluate the quality of Stanbol enhancer annotations

Input text

Paris Hilton is now in New York, after having spent some time in Geneva, Ohio.

Stanbol annotations

```
Paris, France
Hilton Hotel Paris
Paris Hilton
New York State
New York City
Geneva, Switzerland
Ohio state
...
```

© 2011 Adobe Systems Incorporated. All Rights Reserved.

Input

Paris Hilton is now in New York, after having spent some time in Geneva, Ohio.

Desired annotations

```
Paris Hilton
Geneva, Ohio
Ohio State
(
  New York City
  New York State
)?
```

Unwanted annotations

```
Paris, France
Hilton Hotel
Geneva, Switzerland
```

© 2011 Adobe Systems Incorporated. All Rights Reserved.

Running the benchmarks in Stanbol

Input form

apache stanbol™ Apache Stanbol Enhancer Benchmarks - Benchmark Input

- /benchmark/default.txt
- /benchmark/jamaica.txt

Valider

Benchmark example: /jamaica.txt

```
= INPUT =
Bob Marley was born in Kingston, Jamaica.

= EXPECT =
Description: Kingston must be found
http://fise.iks-project.eu/ontology/entity-reference URI
http://dbpedia.org/resource/Kingston%2C_Jamaica

Description: Bob Marley must be found as a musical artist
http://fise.iks-project.eu/ontology/entity-type URI http://dbpedia.org/ontology/MusicalArtist
http://fise.iks-project.eu/ontology/entity-ref

= COMPLAIN =
Description: Miles Davis must not be found
http://fise.iks-project.eu/ontology/entity-typ
http://fise.iks-project.eu/ontology/entity-ref
```

apache stanbol™

Results report

apache stanbol™ Apache Stanbol Enhancer Benchmarks - Benchmark Results

Benchmark input:

Bob Marley was born in Kingston, Jamaica.

Kingston must be found

SUCCESSFUL

(E) <http://fise.iks-project.eu/ontology/entity-reference> URI <http://dbpedia.org/resource/Kingston%2C_Jamaica>

Bob Marley must be found as a musical artist

SUCCESSFUL

(E) <http://fise.iks-project.eu/ontology/entity-type> URI <http://dbpedia.org/ontology/MusicalArtist>

(E) <http://fise.iks-project.eu/ontology/entity-reference> URI <http://dbpedia.org/resource/Bob_Marley>

Miles Davis must not be found

SUCCESSFUL

(C) <http://fise.iks-project.eu/ontology/entity-type> URI <http://dbpedia.org/ontology/MusicalArtist>

(C) <http://fise.iks-project.eu/ontology/entity-reference> URI <http://dbpedia.org/resource/Miles_Davis>

© 2011 Adobe Systems Incorporated. All Rights Reserved.

IKS Adobe

Benchmark definition language

```
= INPUT =
Bob Marley was born in Kingston, Jamaica.

= EXPECT =
Description: Bob Marley must be found as a musical artist
http://fise.iks-project.eu/ontology/entity-type
   URI http://dbpedia.org/ontology/MusicalArtist
http://fise.iks-project.eu/ontology/entity-reference
   URI http://dbpedia.org/resource/Bob_Marley

= COMPLAIN =
Description: Miles Davis must not be found
http://fise.iks-project.eu/ontology/entity-type
   URI http://dbpedia.org/ontology/MusicalArtist
http://fise.iks-project.eu/ontology/entity-reference
   URI http://dbpedia.org/resource/Miles_Davis
```

Repeat for multiple benchmarks on a single page

© 2011 Adobe Systems Incorporated. All Rights Reserved.

Bob Marley was born in Kingston, Jamaica.

Bob Marley must be found as a musical artist

SUCCESSFUL

(E) http://fise.iks-project.eu/ontology/entity-type URI http://dbpedia.org/ontology/MusicalArtist
(E) http://fise.iks-project.eu/ontology/entity-reference URI http://dbpedia.org/resource/Bob_Marley

Miles Davis must not be found

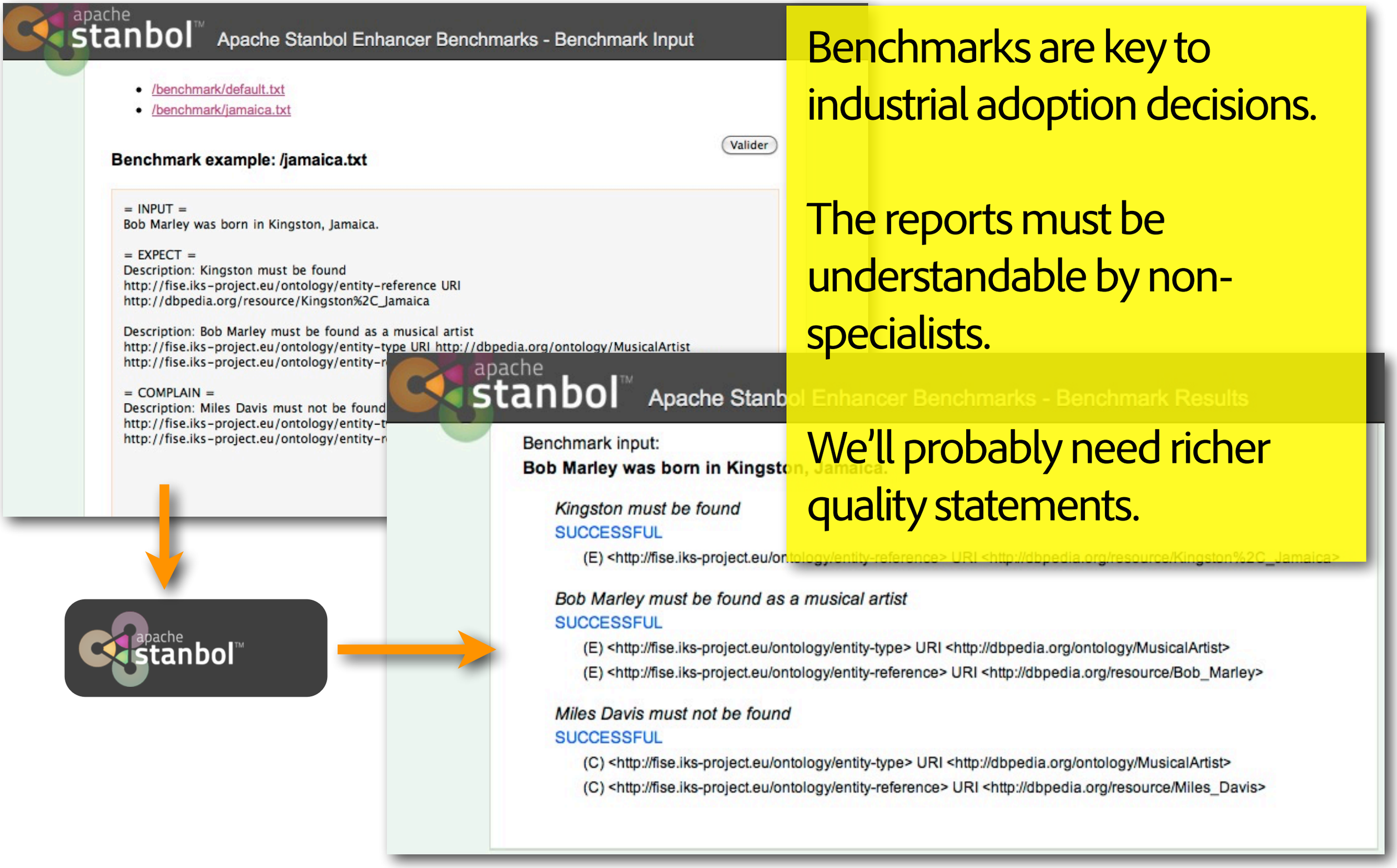
SUCCESSFUL

(C) http://fise.iks-project.eu/ontology/entity-type URI http://dbpedia.org/ontology/MusicalArtist
(C) http://fise.iks-project.eu/ontology/entity-reference URI http://dbpedia.org/resource/Miles_Davis

Very human readable.

© 2011 Adobe Systems Incorporated. All Rights Reserved.

Conclusions



Benchmarks are key to industrial adoption decisions.

The reports must be understandable by non-specialists.

We'll probably need richer quality statements.

© 2011 Adobe Systems Incorporated. All Rights Reserved.

Adobe