Advice:
The time spent on pre-launch analysis is worth the effort to avoid starting from scratch and further alienating already frustrated users by implementing a search which appears to have no connection to what existed before. The analysis provided a baseline from which to measure search success and provide real data from which to make decisions.

5. Impact and Benefits
Business benefits are as yet to be determined. The prospective business impact is that employees will spend less time searching for information and understand the value of categorizing content in order to group content and understand relationships between concepts to make new information discoveries. Extracting more value from existing information is critical, particularly when an industry faces a downturn.

It's hard to say if the benefits could have come from other sources. However, the lack of trust and use of the search tools as they were showed that any improvement would be a success. We will know that our effort made a difference as we monitor the rate of search use and successful and unsuccessful searches. We hope to extend this analysis to determine connections between information retrieved and actual business decisions and their outcomes.

Reasons for success:
The need for an improvement in search bordered on desperation. Many employees have been asking for improved search. The delivery of an improved search tool and user experience coupled with the automatic application of standardized metadata will not only provide a better search experience but will also bolster the design of search-based applications to improve execution in specific use cases.

6. Next Steps
Our next steps are to continue to enhance search functionality with improvements in the search user experience and additional search features. In addition, we will actively pursue the development of search-based applications within SharePoint and connections to external information in other systems and file shares. Expanded search consolidation will also broaden the use of classification of content with standard metadata from the FMCTI Taxonomy. Finally, continued analysis of search terms and response to user requests for content classification will aid in the further development of the FMCTI Taxonomy and lead to improved content classification.

3) CHARLIE HULL

Implementing open source search for a major specialist recruiting firm

1. About the Case Organization
Founded in 1960, Reed Specialist Recruitment is a specialist provider of permanent, contract, temporary and outsourced recruitment solutions, and IT and HR consulting. Reed operates in Europe, the Middle East and Asia Pacific and has more than 3,000 permanent employees working out of 350 offices worldwide across 30 specialisms.

2. About the Challenge
The main objective was to replace the existing recruitment search function which used an Oracle database: searches took several minutes to complete and returned an un-ranked list of results, which staff would then have to work through by hand. The search function was used by all Reed's recruitment consultants to find candidates, jobs and companies and its slow speed and varying quality led to an increased workload.

3. What We Did
Flax designed a flexible and powerful search infrastructure based on the open source search engine Apache Lucene/Solr. A custom indexer written in Java navigated the highly complex legacy database structure used by Reed. Using Apache Tika, plain text was extracted from CVs (Resumes) in Microsoft Office & PDF formats and in many languages including English, Polish and Chinese. A test framework was built to gauge and tune performance. A single search server provides the entire search capability over 12 million records, with sub-second search response times.

4. Challenges and Lessons Learned
The complexity of the existing data model led to a highly flexible pipeline being required, to allow us to easily change how data was mapped into the search index. The new search engine was also such a radical change in terms of performance and features that users required significant retraining to take advantage of these – they simply had forgotten to trust search. Organizations planning this kind of migration should focus on the user experience in particular. Moving to an open source platform may also

require a change in development culture and thus training.

5. Impact and Benefits
The project was delivered on time and on budget – unusual for a project of this scale – and the project was hailed as a great success by the client, part of a large-scale reorganisation of their IT capability. Reed have continued to innovate in search and have developed many further systems using a similar platform. The project was a success due to careful planning, tight project management and the clear benefits of moving to an open source solution – flexibility, scalability, no vendor lock-in and no license fees.

6. Next Steps
Flax has continued to work with Reed to develop powerful search-based solutions for other areas of their business. Flax has also since worked with other recruitment firms to develop similar capabilities.

4) DAVE CLARKE

OASIS: Constructing knowledgebases around high resolution images using ontologies and Linked Data

1. About the Case Organization
Synaptica produces software solutions for: building and managing taxonomies and crosswalks; designing and deploying knowledge organization systems; indexing and enriching content; and optimizing search, navigation and discovery.

2. About the Challenge
Visual images provide a valuable complement to textual information, but a vast amount of the information inside photographs, paintings, diagrams and drawings can be seen but not searched - it has been inaccessible to traditional query methods.

Many business applications could benefit from the ability to search inside images including: medical and scientific imagery, reconnaissance and intelligence, engineering and design, forensics and security, education and cultural heritage.

3. What We Did
Synaptica built a software system called OASIS that allows points and regions inside images to be highlighted and annotated. These visual features are then tagged using taxonomies and Knowledge Organization Systems. The software makes visual content searchable with pin point accuracy. It also promotes knowledge discovery as the application dynamically identifies features and related concepts as the user freely pans and zooms around an image.

The key technologies behind the solution are Linked Data and RDF graph databases. These allow users to connect to vast amounts of high-quality structured information in the Linked Open Data cloud, including authoritative Knowledge Organization Systems and ontologies. The extensive use of ISO and W3C standards and specifications ensures data portability and systems interoperability.

The fusion of several core technologies (high definition imagery, Linked Data, Knowledge Organisation Systems and semantic annotation) represents an innovative solution that opens up new opportunities for enriching visual content.

4. Challenges and Lessons Learned
Working with external data sources from the Linked Open Data cloud presents a number of challenges: (i) external data can be accessed by live queries to remote third-party servers, but these remote systems may not be able to provide adequate uptime availability or responsiveness; (ii) copies of external data can be ingested into local systems, but some datasets, such as DBPedia, may be too large to be accommodated on the available data storage; (iii) while graph databases out-perform relational databases at pattern-based queries, relational databases typically out-perform graph-databases at indexed or full text searches.

Synaptica responded to these challenges by building a flexible system that can simultaneously query data from any number of internal or external data stores. Low volatility data of a manageable size can be ingested while high-volatility or very large datasets can be accessed on remote servers in real-time.

5. Impact and Benefits
The result of the effort is a robust and scalable general purpose toolset that can be used to build taxonomies, access external Linked Data, and annotate image content. By leveraging Linked Open Data, much of which is available without license fees, the time and cost to deploy knowledge organization systems can be greatly reduced.

Using Open Source Search for Recruitment

Charlie Hull - Managing Director
8th June 2015
IKO Conference, Singapore

charlie@flax.co.uk
www.flax.co.uk/blog
+44 (0) 8700 118334
Twitter: @FlaxSearch

What I'll cover today

- The Project & The Solution
- How we did it
  - **A flexible pipeline in two parts**
  - **Transforming the UI**
  - **Performance**
  - **Issues**
  - **Results & benefits**

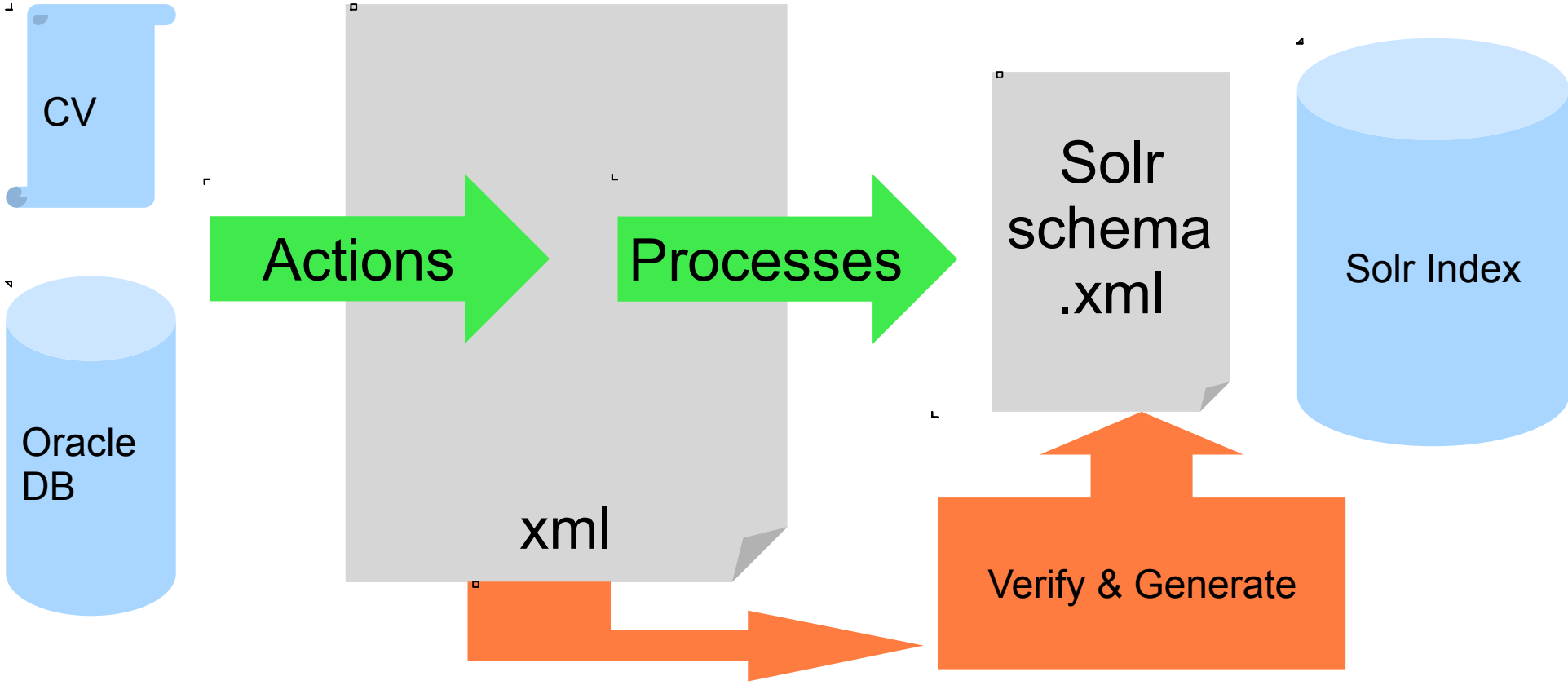
The Project

- The client: *Reed Specialist Recruitment*
- The data
  - **Tens of millions of items to search**
  - **Hundreds of fields in the database schema (which will change in the future)**
  - **CVs (resumés) in Word, PDF formats**
  - **Multiple languages (English, Polish, Chinese...)**
- The problem
  - **Search takes several minutes**
  - **3000+ users familiar with the old system**
  - **No foundation for innovation**

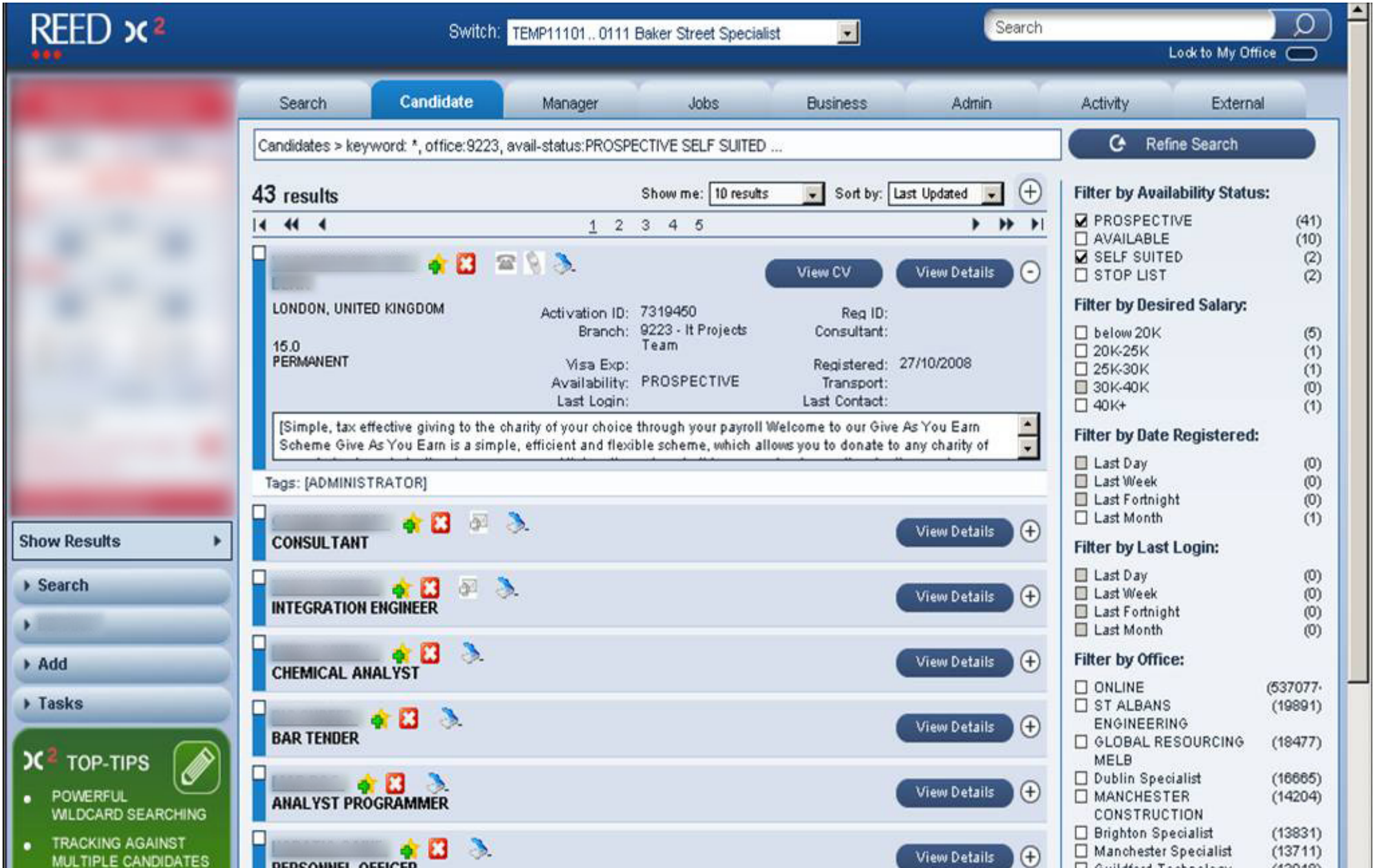
The Solution – Apache Solr

- Flexible and extendable
  - **This is only the first wave of development**
  - **A need for complex business rules to drive the search – Boosts & FunctionQueries**
- Economically scalable
  - **Much more data to come**
  - **Too hard to predict future cost of commercial, closed source alternatives**
- Support available - from  & 

The Indexer & The Config Tool



Transforming the UI



Results & benefits

- Project delivered on time and under budget
- Now live across 350 offices UK & worldwide
- 24/7/365 support provided by Lucidworks & Flax
  - Since downgraded as very few issues
- A very happy client!