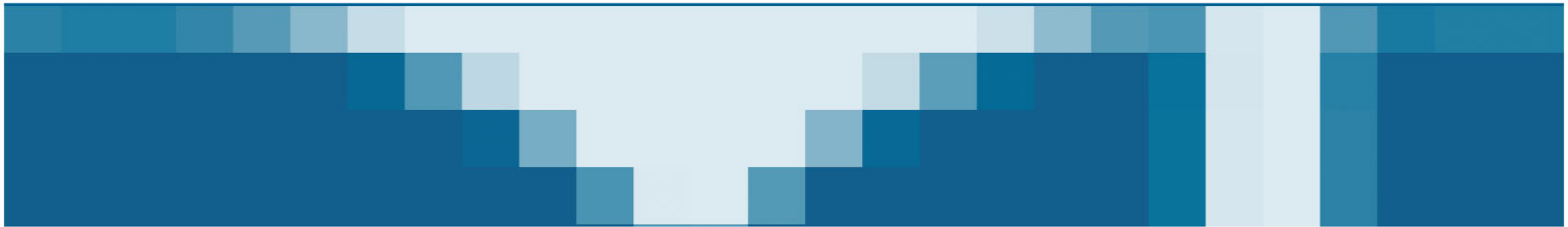


## MATRIXWARE'S ALEXANDRIA PROJECT



## OUR MUSE



From the early 3rd century BC and for nearly 800 years Alexandria “became the matrix for a new spirit of critical inquiry, for a heightened perception of knowledge as a tool and the quest for knowledge as a collaborative process.”

(Derek Flower, *The Shores of Wisdom*)

## OUR VISION



Matrixware's Alexandria provides an extensible, global storage facility for high quality scientific, technical, and business information. Initially populated with patent documentation, Alexandria supports a global R&D community by delivering an infrastructure capable of exposing the diversity of data services required by professional researchers, information specialists, and software engineers.

## SHORT TERM GOALS



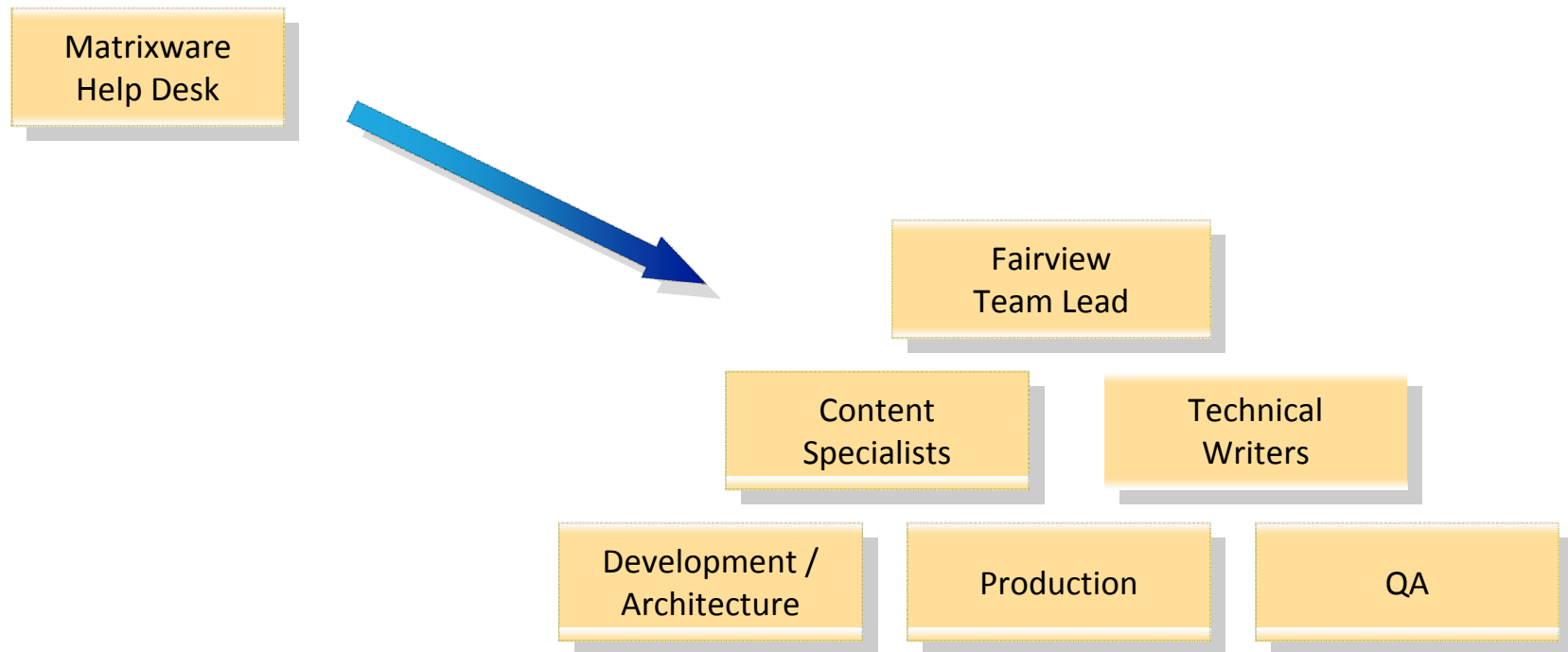
- Develop a high quality, complete archive of international patent documentation
- Deliver patent data as a commercially viable data feed
- Provide flexible, easy-to-use data services to Information Retrieval (IR) scientists in order to support ongoing academic research
- Give users the ability to easily extend the archive content
- Design an infrastructure and framework that allows the integration of annotation frameworks such as GATE from the University of Sheffield
- Expose resulting annotations to research community
- Design, build, and maintain a number of searchable indexes of the archive using technologies such as Lucene, Lemur/INDRI, Terrier, etc
- Explore alternatives to conventional storage technologies like MonetDB, Intersystems Cache, and others

## OUR COMMUNITY



- IRF membership
- Sponsored researchers
- Leonardo application developers and users
- Commercial partners
- Matrixware's professional services and product development teams
- Matrixware's customers

# THE ALEXANDRIA ORGANIZATION

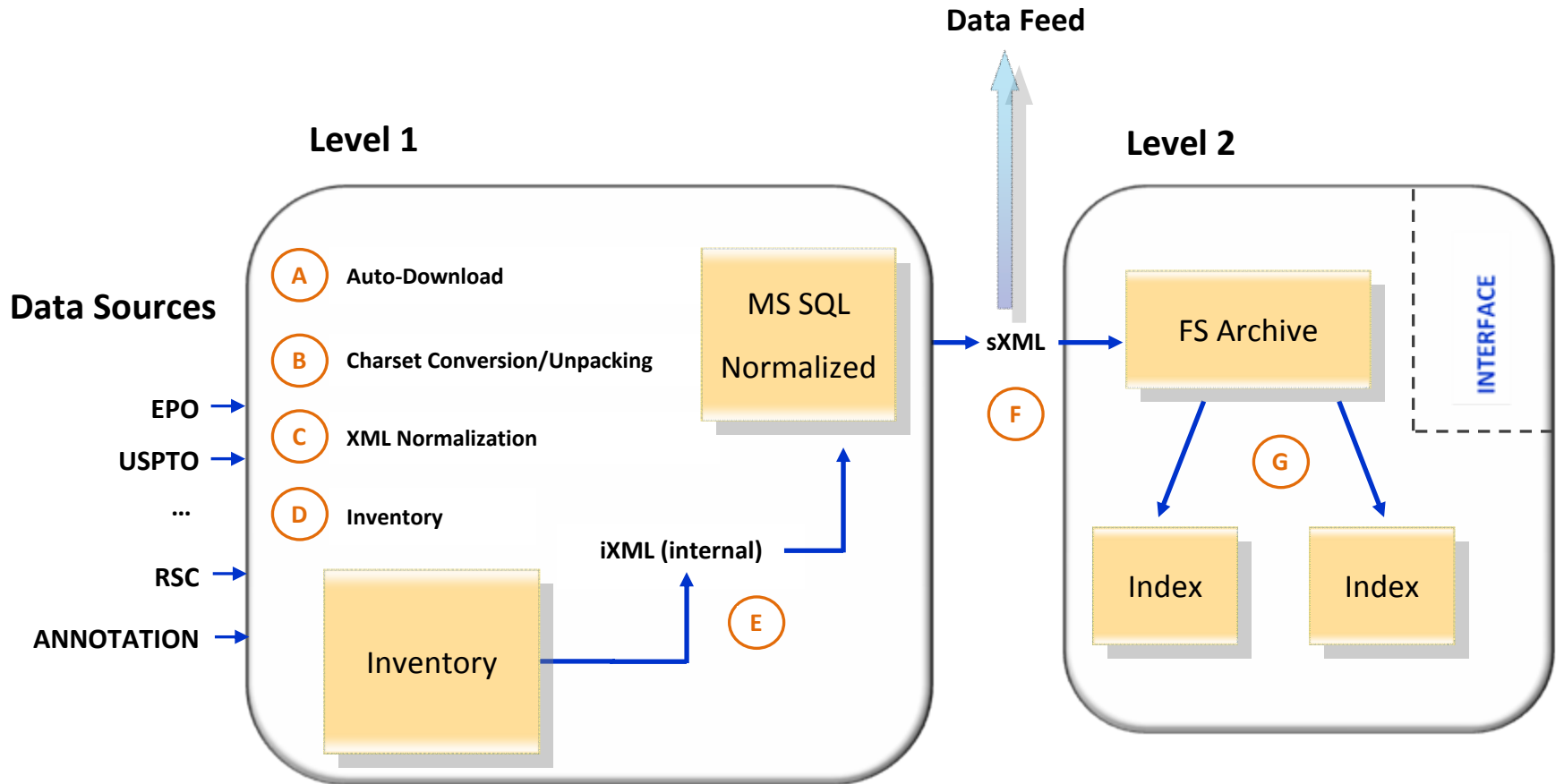


## THE MATRIXWARE – FAIRVIEW RESEARCH RELATIONSHIP

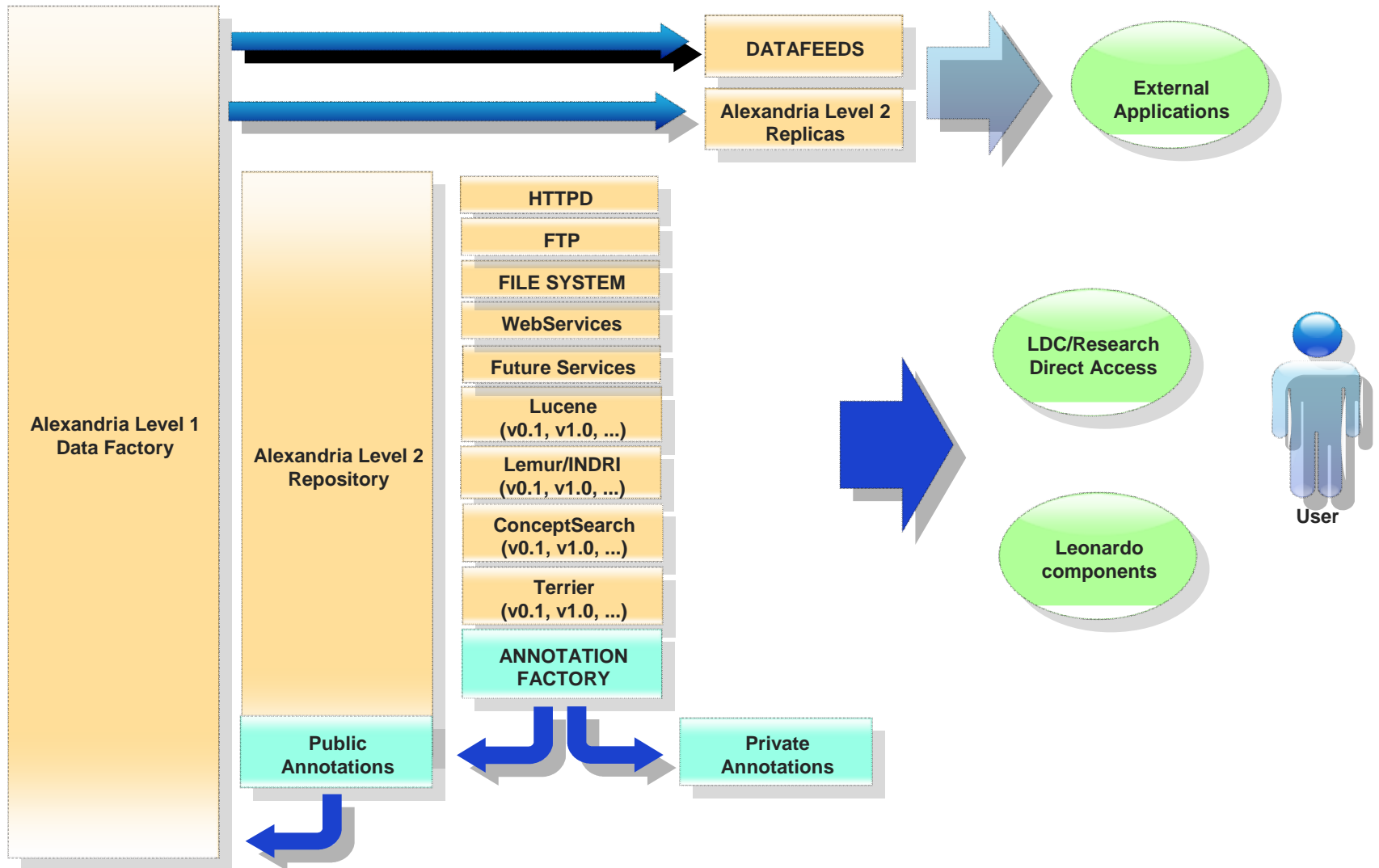


- Matrixware owns and operates the Alexandria archive.
- As a service provider, Fairview Research provides Matrixware with access to an experienced team of content specialists and data architects.
- As commercial partners, Matrixware and Fairview Research are the first of many relationships that will form the collaborative marketplace supported by both Leonardo and Alexandria.

# THE ALEXANDRIA DATA FACTORY



# THE ALEXANDRIA ARCHITECTURE



## CURRENT ALEXANDRIA CONTENT HIGHLIGHTS



- A single, uniform XML format created from merged data sources
- Contains multiple name formats (e.g., original, standardized) for search purposes
- Is multilingual with language tags
- Logs changes and stores update information
- Provides a rich, comprehensive data set
  - > Covers PCT Minimum documentation
  - > Retains every publication
  - > Includes legal status and citations

## DOCUMENT ARCHIVE FORMAT

---

The Level 2 document archive consists of XML files stored in a predictable path layout based on document's country code and publication number.

- Example: publication AT-334199-B

Current archive root=<http://alda.matrixware.com>

Two Character Country Code + 12-character, 0-padded document number

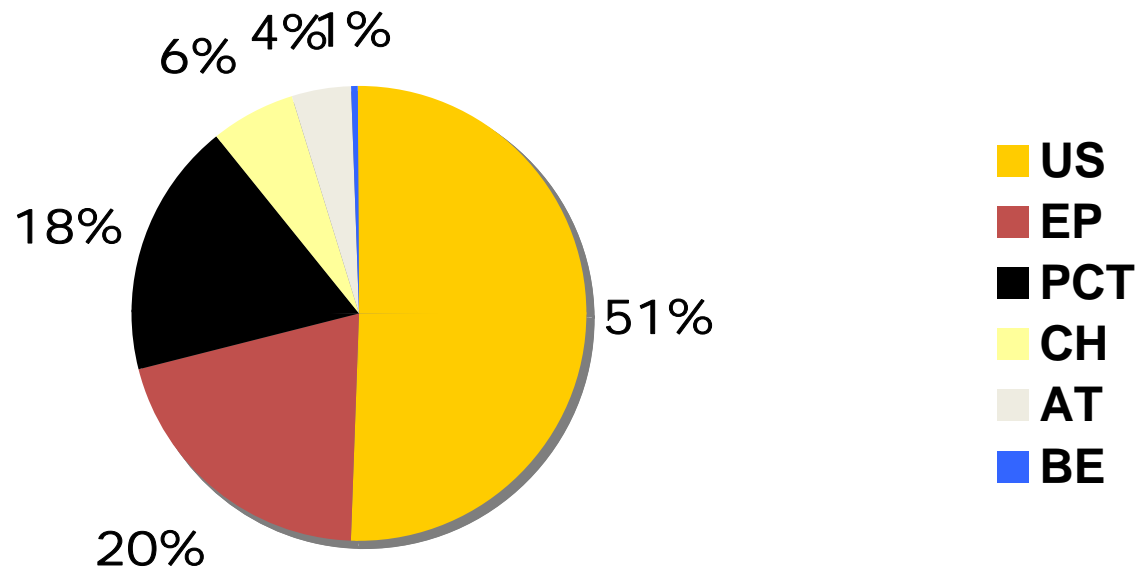
`$root/AT/000000/33/41/99/AT-334199-B.xml`

or

`http://alda.matrixware.com/AT/000000/33/41/99/AT-334199-B.xml`

## COLLECTION AS OF NOV 2008

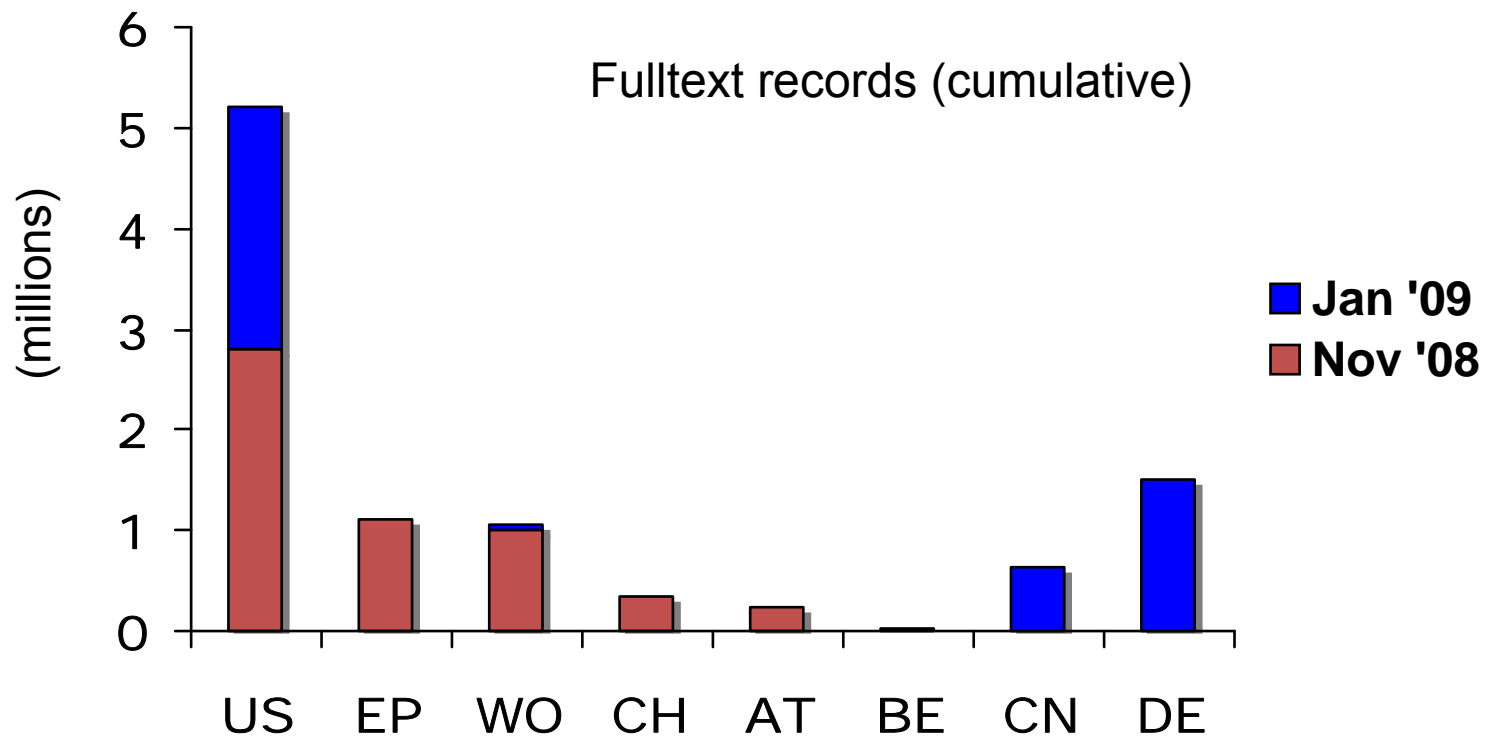
Currently Alexandria contains roughly 67 million patent records from more than 70 countries, including 5.5 million fulltext records.



Fulltext records

## COLLECTION AS OF JAN 2009

Alexandria will contain roughly 67 million patent records from more than 70 countries, including 9.5 million fulltext records.



## WHAT'S NEW IN 2009?



- Continue adding to existing collections such as additional Chinese, PCT, Japanese data
- Develop new content focusing on acquiring additional national data from Asia, Europe, and the Americas
- Continue developing aggressive data QA procedures to track, catalog, and correct systematic data problems
- Promote more level 1 data to level 2 such as images, DNA sequences, chemical data, etc
- Enhance and expand indexing services and related Leonardo connectors
- Integrating annotated data from sources such as the SAM2 project
- Acquire and integrate other non-patent collections such as SEC data, case law, open source scientific publications, and other 3rd party sources
- Acquire and integrate dictionaries, thesauri, ontologies, etc

## A FEW ANTICIPATED QUESTIONS



- Is there a DTD for the standard archive? Why is it different from the DTD's used by the patent offices?
- What is a patent family and what relevant information is available?
- What languages are represented, and how can I find matching text in different languages?
- What kinds of updates are applied to the patent data, and how are they tracked?
- What exactly are priority-claims?
- How do I get access to the data?
- How do I traverse the entire Alexandria archive?
- How do I select a subset of the Alexandria archive?

## DOCUMENTATION AND SUPPORT



- An Alexandria FAQ is available today.
- We are expanding our available documentation to include detailed information on Data Sources, Coverage, Content, and Architecture.
- We are also building a collection of modern training materials to support both IP and IR users.
- To ask any questions about Alexandria and to submit your feedback please contact:

**[Alexandria.support@matrixware.com](mailto:Alexandria.support@matrixware.com)**

**THANK YOU ▪**



For more information, visit the Alexandria booth in the exhibit hall.