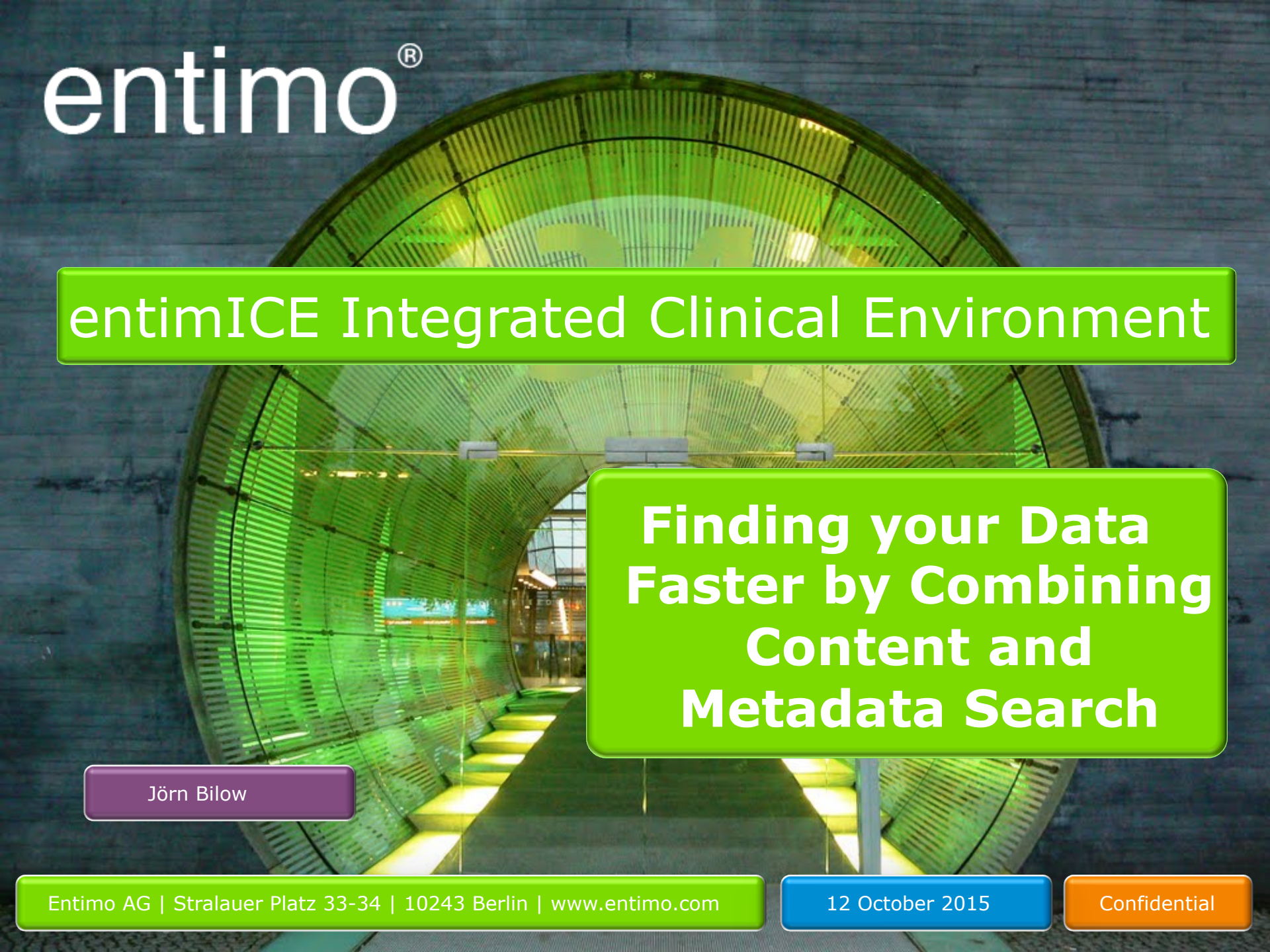




entimo[®]

entimICE Integrated Clinical Environment

**Finding your Data
Faster by Combining
Content and
Metadata Search**

Jörn Bilow

-

Searching Clinical Data

Search

Search

www.Online-Shopping.com

You recently searched for:

You may also be interested in:



There were 57 serious adverse events in studies with compound XYZ.

- **Regulatory compliance:**
 - Access rights
 - Blinding rules
 - External users
- **Information in multiple formats:**
 - PDF, DB Tables, CSV, SAS Datasets
 - Metadata
 - CDISC Attributes
- **Performance:**
 - Speed
 - Comprehensive reports
 - Fast access to relevant results

Indexing Server

- **Based on solid technologies:**

- Apache Lucene
- Apache Tika
- Akka
- Xtext



■ **Indexed information:**

- Object metadata
- Object attributes (e.g. name, comment, creation date)
- Object historic lists (status history, user change history)
- Object version attributes (e.g. last modification date or last change user)
- File content (e.g. text, PDF, MS Office etc.)
- Content of programs, scripts, log files, listing files
- Program execution trace
- Content of datasets and tables (Database tables, SAS datasets, CSV datasets)
- Mappings

- Highly parallel indexing
- Various extractors (SAS, CSV, PDF, metadata) work in parallel
- Different index domains written in parallel
- No client impact as background re-indexing on the server
- Data stored as “fuzzy” match, metadata as “exact” match

- Configuration options:
 - Custom extractors (possible even with any SAS program!)
 - Include/exclude sets
 - Multiple remote import directories
- Remote import dialog supported:
 - Check before import
- All entimICE metadata included:
 - Allows to combine data and metadata in queries for precise searches

- Alternative to Lucene native language
- Searching with familiar terms
- CDISC attributes, metadata
- Search can include:
 - Qualified keys, wildcards, date & time stamps
 - Information about users, versions, audit trail, linking
 - Directory tree & hierarchy information

Examples / Demo

- Scenario 1: **Content search**
- Scenario 2: **Object history**
- Scenario 3: **Object metadata**

- Scenario 4: **Object metadata, multiple search criteria**
- Scenario 5: **Combining content and metadata search**

- User searches for objects
 - That have (not) been updated in a specific timeframe
- User searches for objects
 - That are linked to a specific dataset
- User searches for outputs
 - That were created by program XYZ within a certain timeframe...

Conclusion / Outlook

- Efficient search engine
- Combined content and metadata information
- Results reflect individual user access rights
- Domain Specific Language with familiar terms

- **More scalability options:**
 - Multiple machine setups, multi-index setup (sharding)
- **Smarter indexing:**
 - Special parsers for code, log files, ODM and other file types
 - More options for dataset extraction (typed indexing)
- **Integration:**
 - Combination with other search engines' results

Questions and Answers

