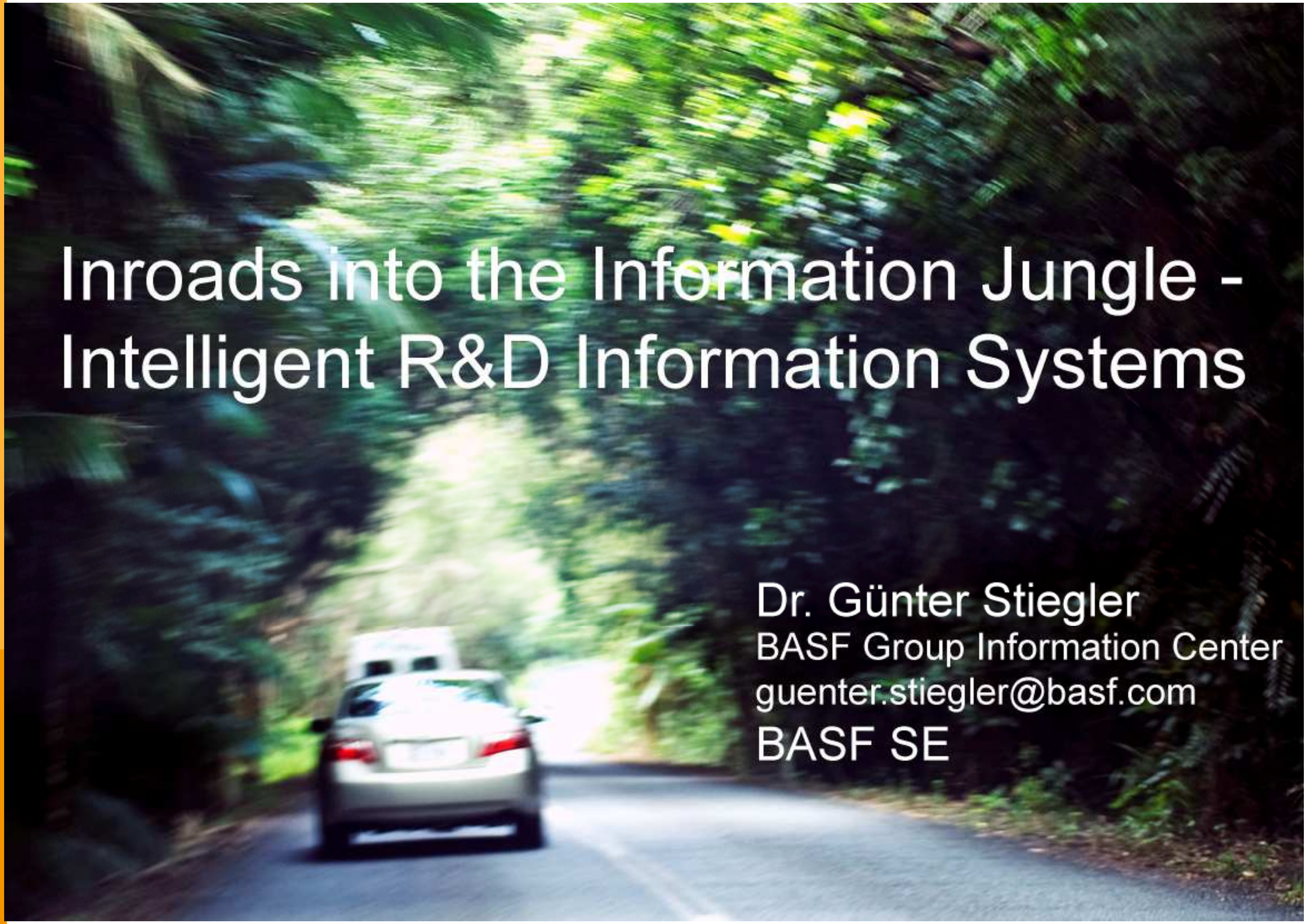


Inroads into the Information Jungle - Intelligent R&D Information Systems

Dr. Günter Stiegler
BASF Group Information Center
guenter.stiegler@basf.com
BASF SE

BASF
The Chemical Company

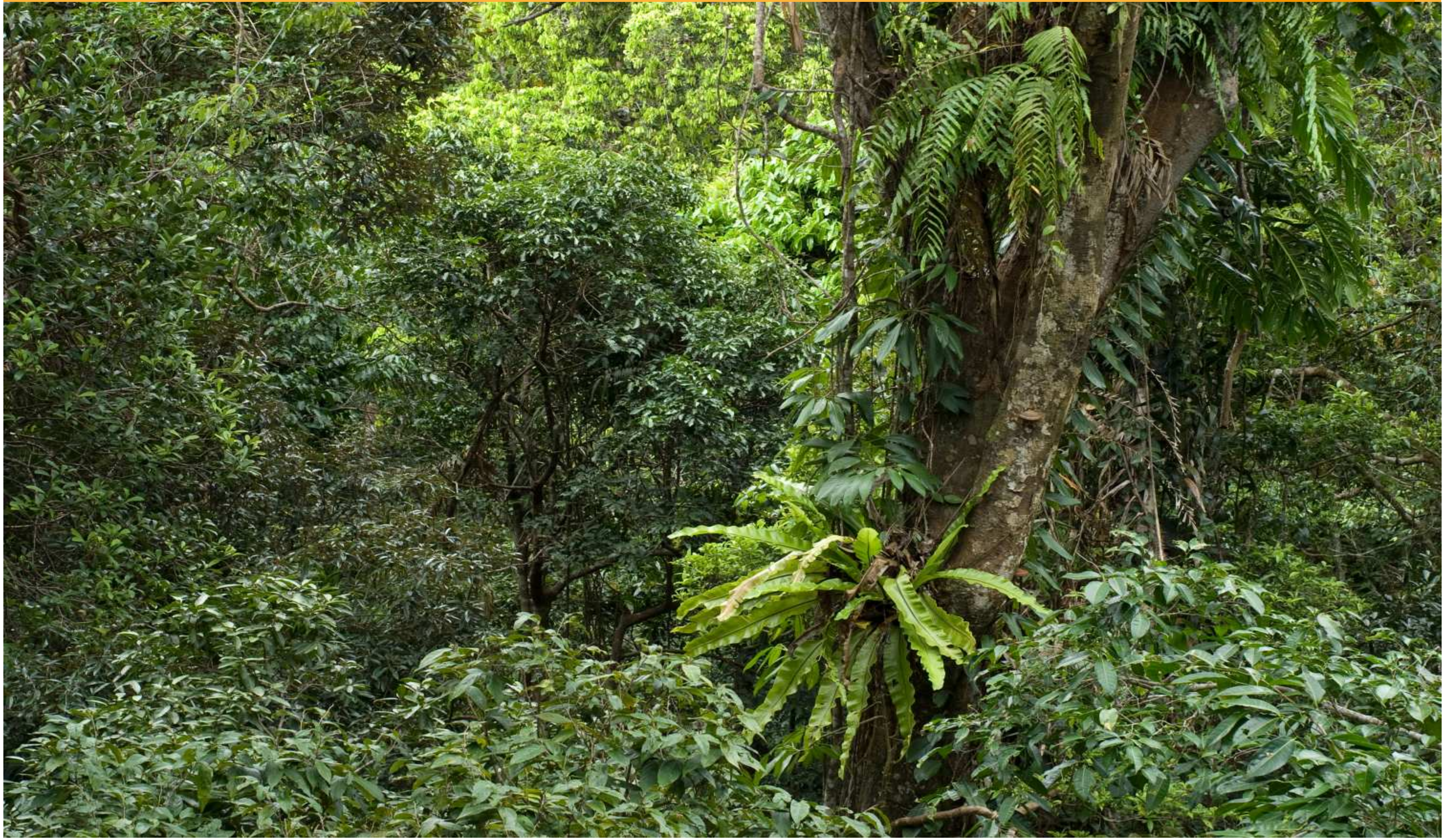
ICIC Nice, October 20, 2008

Inroads into the Information Jungle: Agenda

The Chemical Company

1. **Introduction: The Information Jungle**
2. BASF Group Information Center
3. Traditional R&D Information Systems
4. Requirements from the Business Process
5. Intelligent R&D Information Systems
6. What`s to be done?

The Jungle



The Context

BASF
The Chemical Company

Inroads into the Information Jungle: Agenda

1. Introduction: The Information Jungle
2. **BASF Group Information Center**
3. Traditional R&D Information Systems
4. Requirements from the Business Process
5. Intelligent R&D Information Systems
6. What`s to be done?

BASF Group Information Center

BASF
The Chemical Company

Our Mission:

- The right information
- At the right time
- In the right context

Our Responsibility:

- Provide Scientific, Technical and Business Information for the BASF Group

Our Portfolio:

- Searches by information professionals
- Enduser information systems
- Access to original literature
- Customized solutions for communities

Inroads into the Information Jungle: Agenda

BASF
The Chemical Company

1. Introduction: The Information Jungle
2. BASF Group Information Center
3. **Traditional R&D Information Systems**
4. Requirements from the Business Process
5. Intelligent R&D Information Systems
6. What`s to be done?

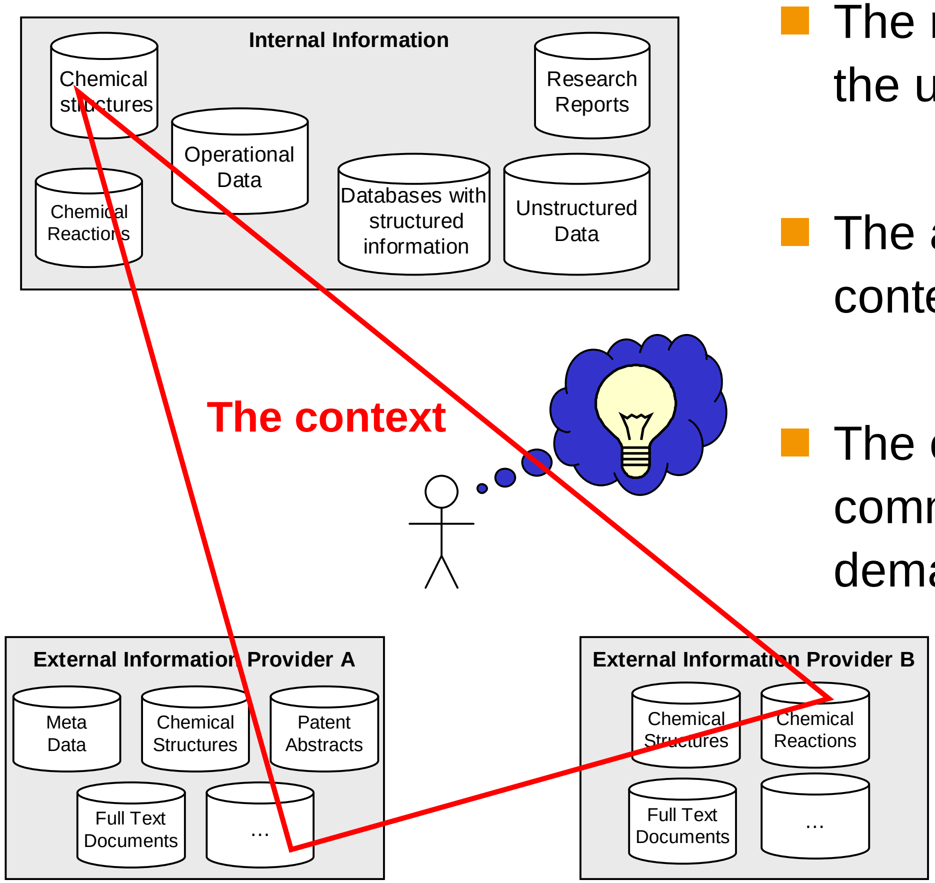
Traditional R&D Information Systems

- The internal business process, the type of content or commercial aspects of external providers determine the design
- Each system has its own look and feel and its own query language
- Typical questions in the R&D process need to be answered across these boundaries and this is **not** supported by the traditional information systems

→ **Traditional inhouse information systems store different types of information in different silos.**

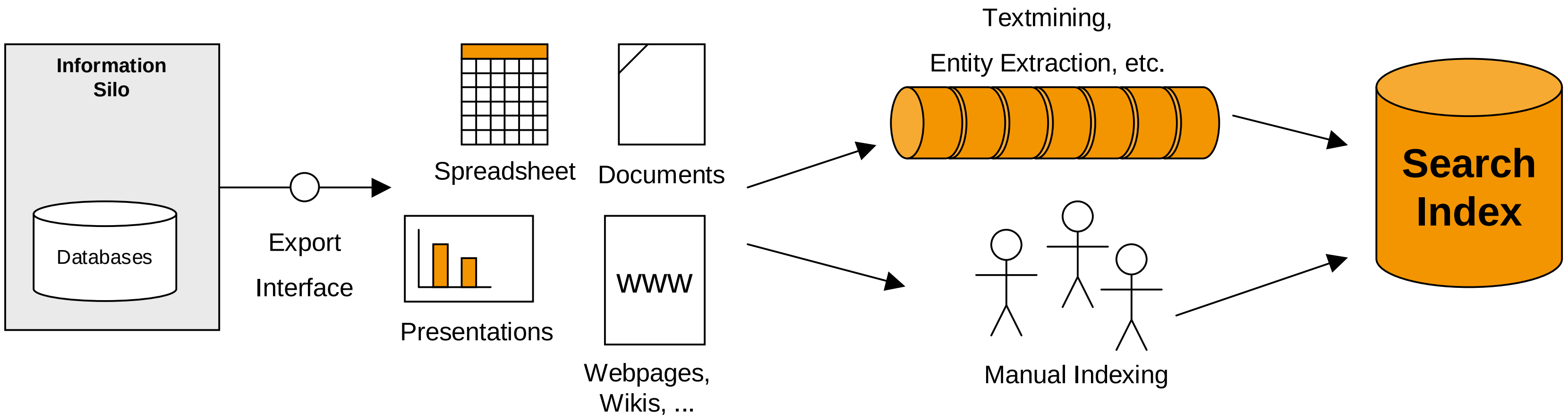
→ **They are inflexible for searching for information in the desired context.**

Information Silos: Consequences



- The relevant information is distributed and the user **must know** where to search for it
- The aggregation of the data in the right context is normally manual work
- The customization to specific needs of a community requires great effort and demands data migration and duplication

Information Silos:
Growth of Unstructured Information



- Export of data to desktop applications to compensate for poor integration and lack of flexibility
- The postprocessing of this extracted data is done in a semi- or unstructured way
- Important and relevant new information is available only in an unstructured form
  → **"The information jungle"**

The consequence:

- **Complicated and expensive processes are necessary in order to make the information searchable in the desired context.**

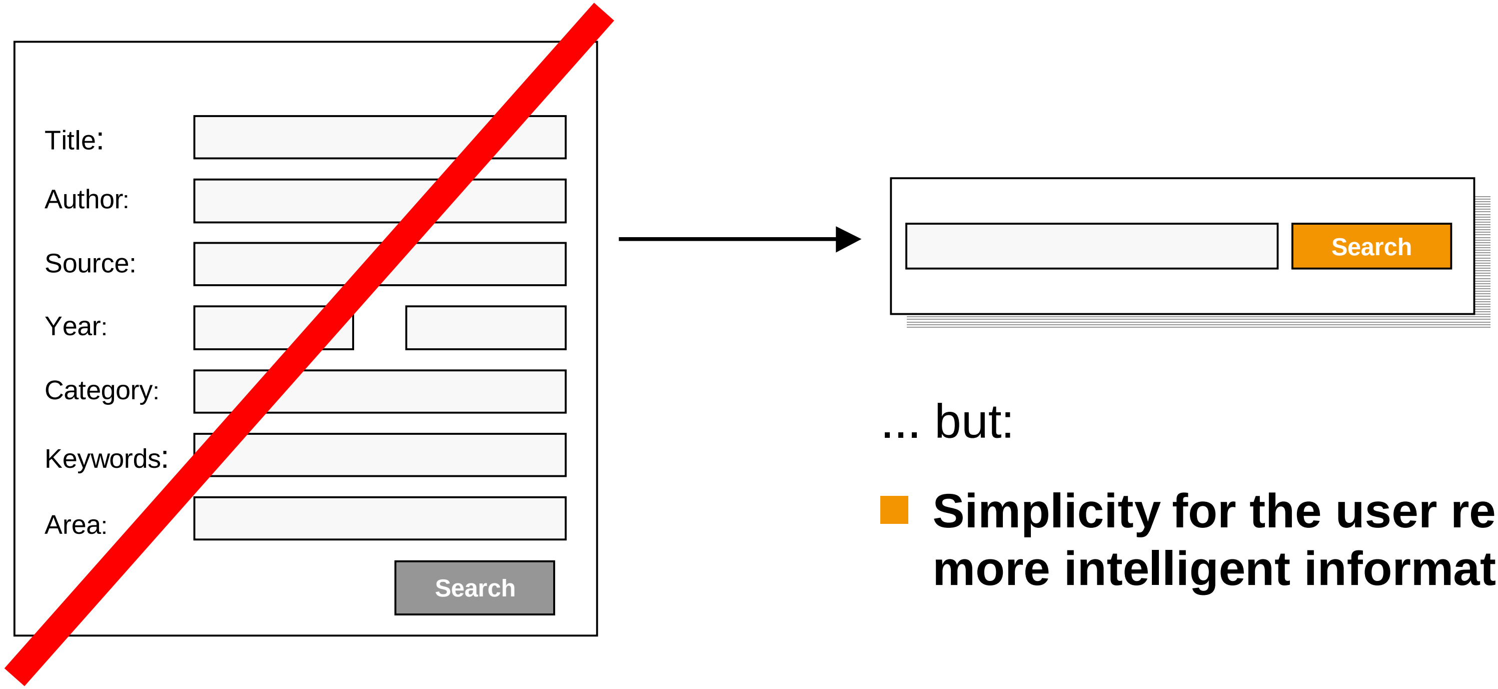
Inroads into the Information Jungle: Agenda

1. Introduction: The Information Jungle
2. BASF Group Information Center
3. Traditional R&D Information Systems
4. **Requirements from the Business Process**
5. Intelligent R&D Information Systems
6. What`s to be done?

Requirements from the Business Process

- Provide an integrated view on relevant internal and external sources
- Break down the boundaries between different types of information:
  - Document type information
  - Structured data
  - Chemical entities
- Provide a "more intelligent search system":
  - Flexible ranking
  - Navigation in the result set
  - Post-processing
- Enhance the information with additional indexing for special purposes
- Build fast and dynamic solutions for new communities or business processes
- Integrate the application into the community's working environment
- Enable intuitive usage (The "Google effect")

The "Google-Effect":
User Demand for a "Simple" System



... but:

- **Simplicity for the user requires a more intelligent information system!**

Analysis of Requirements

- An intuitive and "simple" system for the user requires complex technology and processes for the content preparation
- There is no single commercial provider who can deliver a solution that covers all aspects of
  - Full range of data (patents, chemical entities, business information,…)
  - Functionality
  - Flexibility for different communities
  - Competence for software development
- An intelligent R&D information system requires external content providers
- The development requires powerful software components
- The support of the R&D process with the internal R&D knowledge is an internal core competence
- A clear information architecture and internal governance are essential

Inroads into the Information Jungle: Agenda

BASF
The Chemical Company

1. Introduction: The Information Jungle
2. BASF Group Information Center
3. Traditional R&D Information Systems
4. Requirements from the Business Process
5. **Intelligent R&D Information Systems**
6. What`s to be done?

Architectural Principles

- Keep operational systems for data input separately from search systems
- Integrate different datatypes, for example documents, chemical entities, etc.
- Use available structured information and metadata
- Incorporate semantic methods e.g. text mining iteratively
- Design flexible front-ends for different communities
- Use powerful components for speedy development: configuration instead of programming
- Define clear components and interfaces

New Intelligent Systems: Fundamentals

What Information?

Relevant data from internal and external sources, e.g.

- Synthesis information
- Phys.-chem. data
- Tox. information
- Research reports
- Literature information
- Business data
- …

How to access it?

Technical interfaces (connectors) to the different technical representations of the source information:

- Relational databases
- Document management systems
- Folders and directories
- Chemical structure management systems
- …

How to prepare it for intelligent access?

Flexible combination of methods for content Preparation and indexing:

- Format conversion
- Normalization
- Lemmatization
- Entity recognition
- Categorization
- Additional indexing
- ...

How to process the query for a "simple" search?

Translation of the user‘s query input to the machine‘s search process and language

- Normalization
- Recognition of phrases
- Natural language query support
- Thesauri, Synonyms
- ...

How to present the result in the right context?

- Aggregation of results
- Ranking
- Abstract preparation
- Highlighting
- Visualization
- Clustering
- Navigation
- …

Security und Systems Management

Software Development

New Intelligent Systems: Use of Core Components

- A powerful search technology together with content extraction, analysis, transformation and enrichment with additional information from e.g. text mining

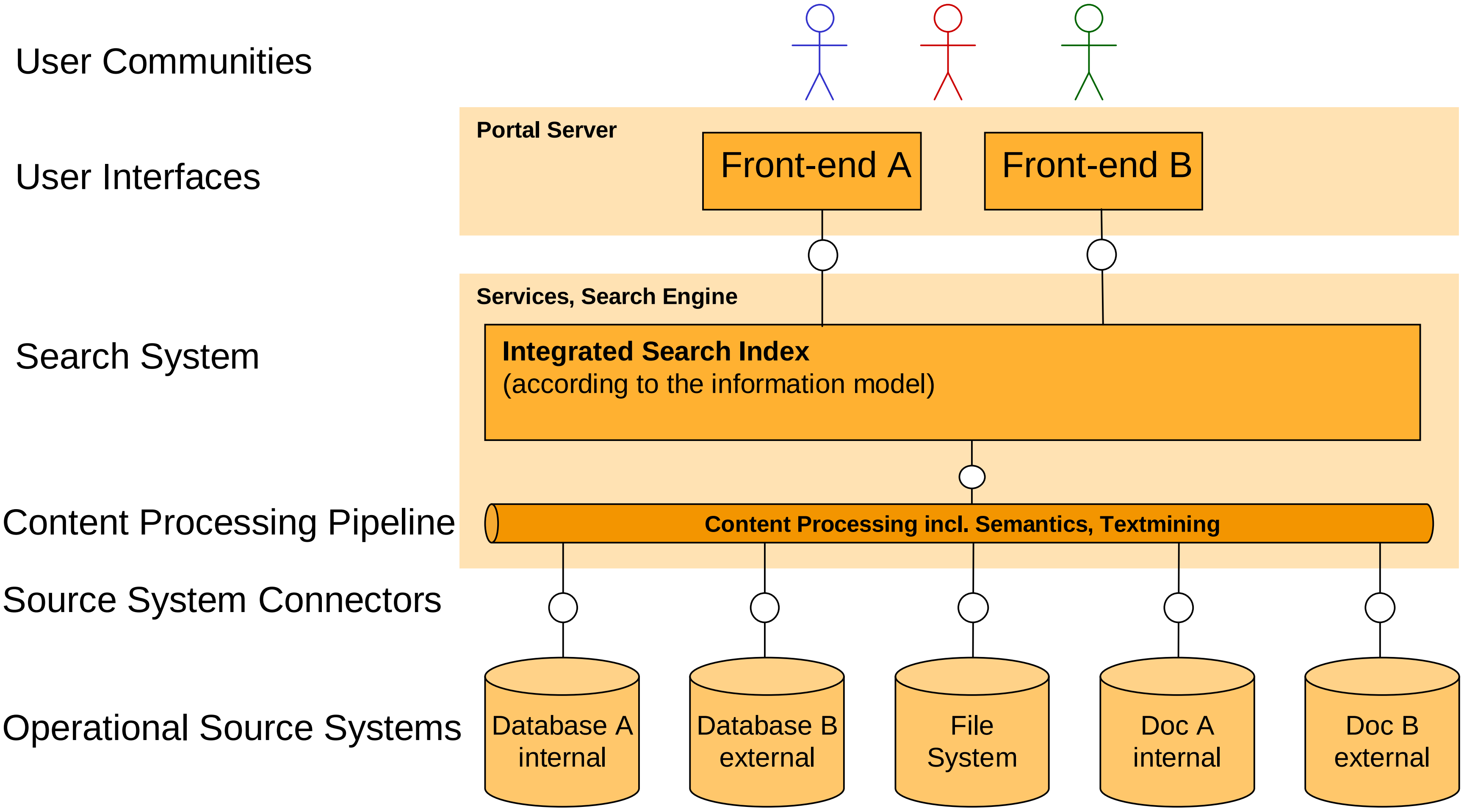
➔ Search engines

- Dynamic front-end components with enough flexibility for different user communities

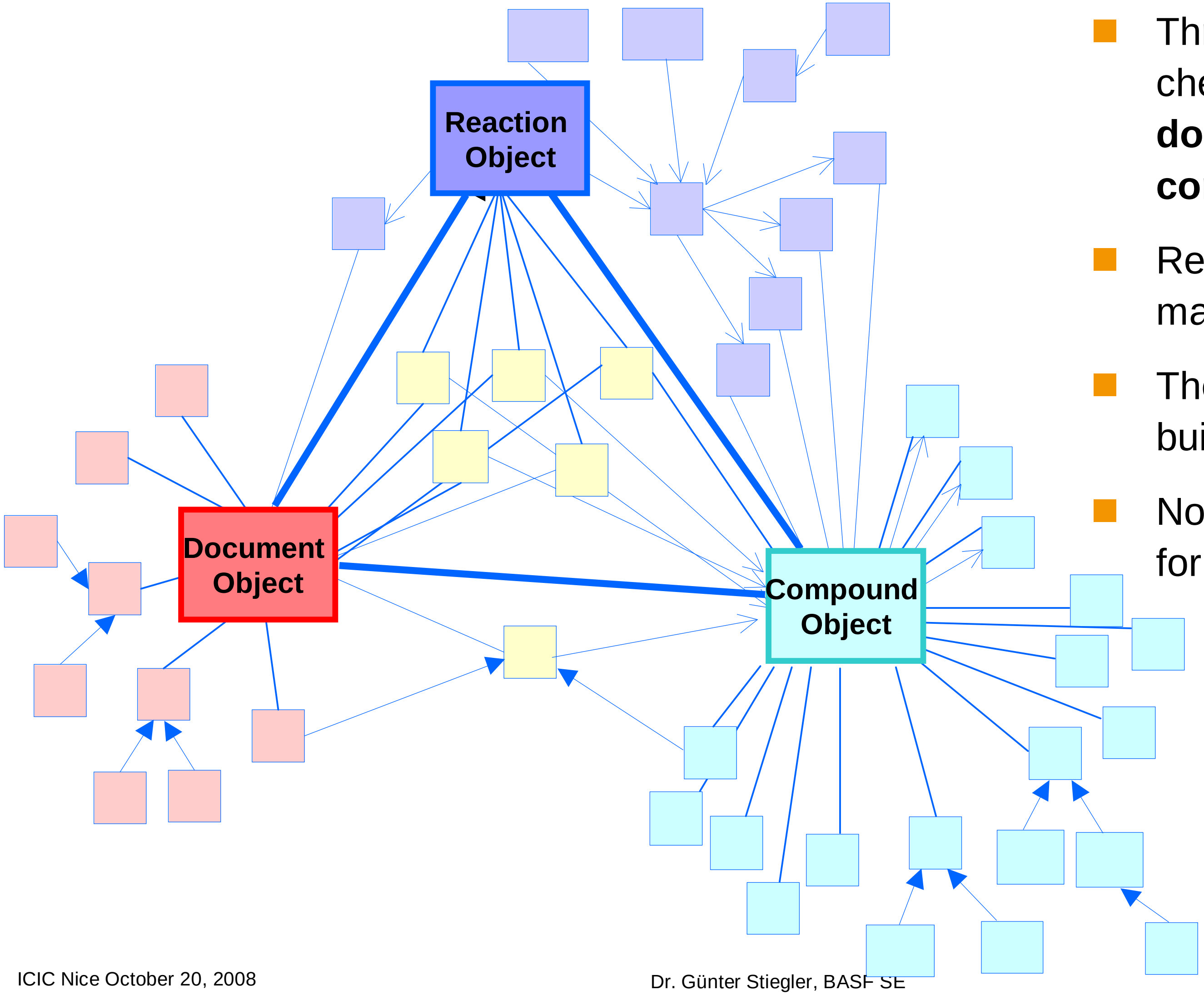
➔ Portal technology

New Intelligent Systems: Architectural Concept (1) - Overview

BASF The Chemical Company



New Intelligent Systems:
Architectural Concept (2) – Information Model



- Three main entities for chemistry R&D information: **document, chemical compound, chemical reaction**
- Relevant source information is mapped to this data structure
- The integrated search index is built on this model
- No deep link to source systems for result browsing

Inroads to the Information Jungle: Agenda

BASF
The Chemical Company

1. Introduction: The Information Jungle
2. BASF Group Information Center
3. Traditional R&D Information Systems
4. Requirements from the Business Process
5. Intelligent R&D Information Systems
6. **What`s to be done?**

What's to be done?

- Improvement and more variety of result ranking
- More dynamic content presentation in the front-end (a dynamic portal application instead of the content manager serves for different contexts and communities).
- More semantics, more text mining
- More information analytics (data-mining, visualization, statistics)
- Information providers to deliver the **content** in a standardized exchange format
- **The most difficult task:** Reduction of the growth of unstructured information

Inroads into the Information Jungle

\\ Photos: Dr. Günter Stiegler \\