FIZ Frankfurt Biotechnology Innovation Center

Frankfurter Innovationszentrum
Biotechnologie

**Applying Machine Learning and
Human Editors to Patent Data**

Dr. Christian Garbe

FIZ Frankfurt Innovation Center

1st Building Complex

June 2002: Foundation

July 2004: Opening

Nov 2002: PPP

Jan 2006: Occupancy rate: 100% 1st CC

2nd Building Complex

Nov 2008: Opening

2012: 3rd CC planned

Jan 2007: Occupancy rate: 80%

July 2009: Occupancy rate: 100% 2nd CC
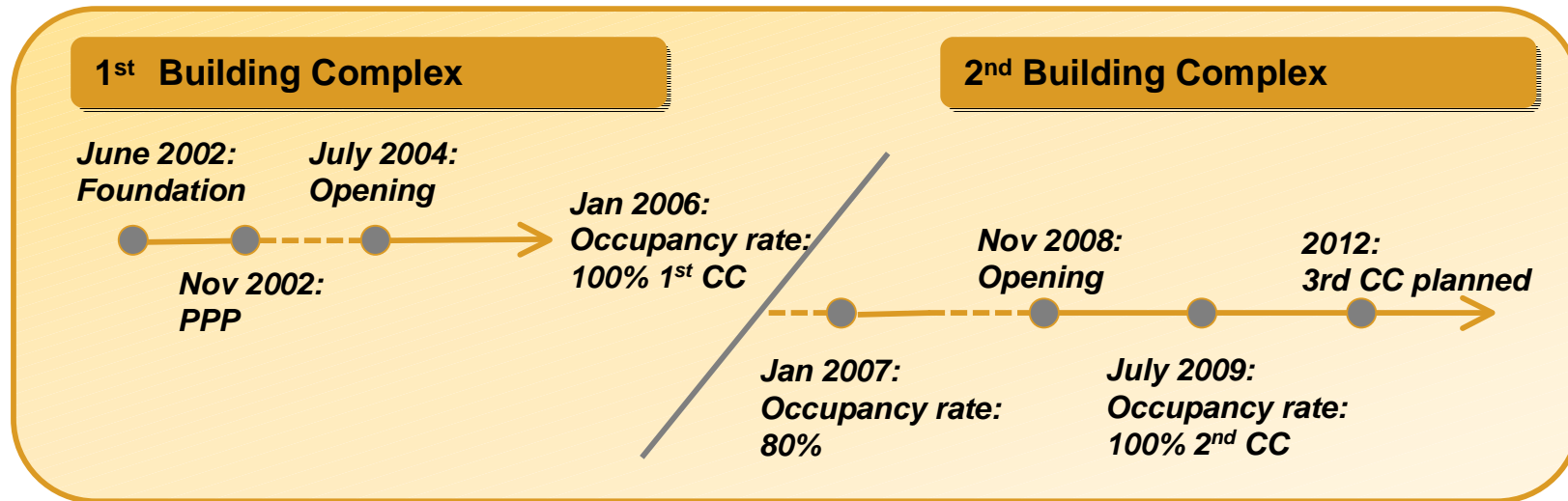
Infrastructure	Investment in Mio EUR
1st CC as combined incubator : 6.500 m² lab and office space	ca. 25,8
2nd CC 8.500 m² commercial lab and office space	ca. 24,2
Conference center	ca. 2
Private investments by tenants	ca. 25
Science City Frankfurt Riedberg	
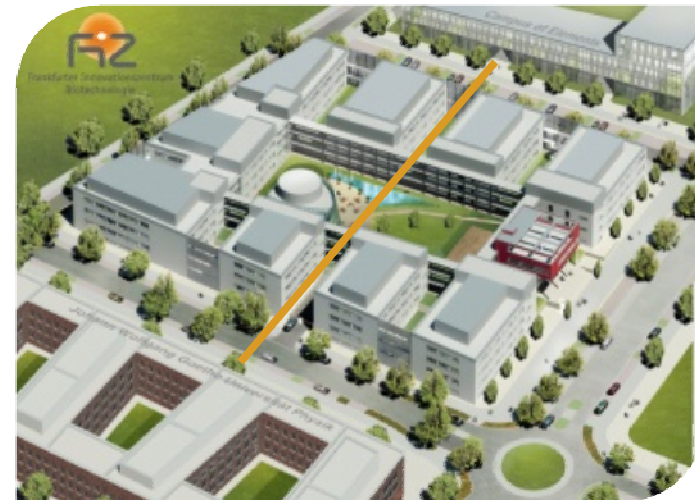
FIZ Frankfurt Innovation Center

FIZ is „the **central contact point**, closing the gap between **basic research** and **commercialization** and thus FIZ supports the **creation** of **added value by turning innovations** into **marketable products**".

Infrastructure

Entrepreneurship

Excellence

Expansion
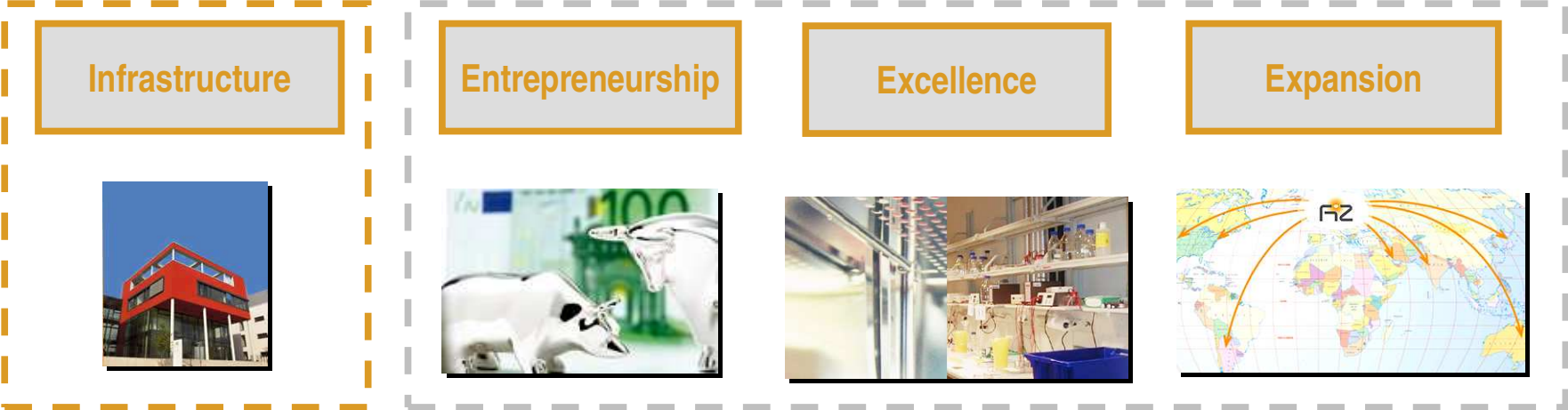
FIZ Frankfurt Innovation Center Infrastructure

Tenants of FIZ benefit from the **geographical location** and

the **modern equipment** of FIZ.

Geographical Location

- ✏ *Proximity* to the international *Airport* Frankfurt

- ✏ *Science City Frankfurt-Riedberg*

- ✏ *Neighborhood* to all **scientific faculties of the** Johann Wolfgang Goethe-University & other **excellent research institutes**

- ✏ Closeness to **international pharmaceutical companies**

Laboratory – and Office Space

- ✏ **Individually designed** and **modern office- and laboratory space (~ 14.000 m² / ~ 150,700 sq ft)**

- ✏ **S1-S3 laboratories, cleanrooms, customized interior design**

- ✏ **FIZ ConferenceLab**

FIZ Frankfurt Innovation Center Entrepreneurship

The **economic focus** and the **network** of FIZ reflect **entrepreneurship**.

Economic Focus

- ☑ **Focus** on **marketability** and **commercialization of innovation**

- ☑ **Scientific and Business Advisory Board**

Network

- ☑ **Heterogeneous structure of tenants** supports **synergies and co-operation**

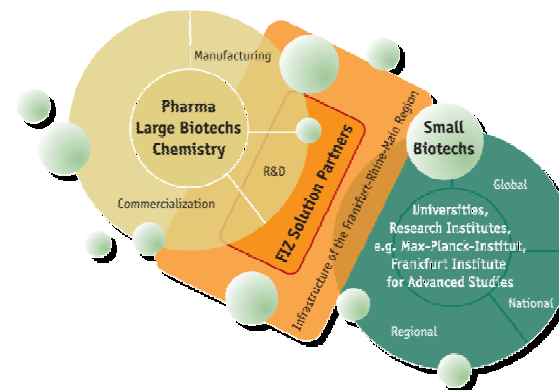# FIZ Frankfurt Innovation Center Excellence

> **Focused research**, a **high degree of networking** and a structured **interface** enable an **efficient technology transfer**.

Research

- **Indication areas at FIZ**
  - Diseases of the Central Nervous System
  - Inflammatory Diseases
  - Proteomics

- **Network** enables access to **the equipment of local public institutes**

Interface

- FIZ as **interface** between **basic research** and the **pharmaceutical industry**

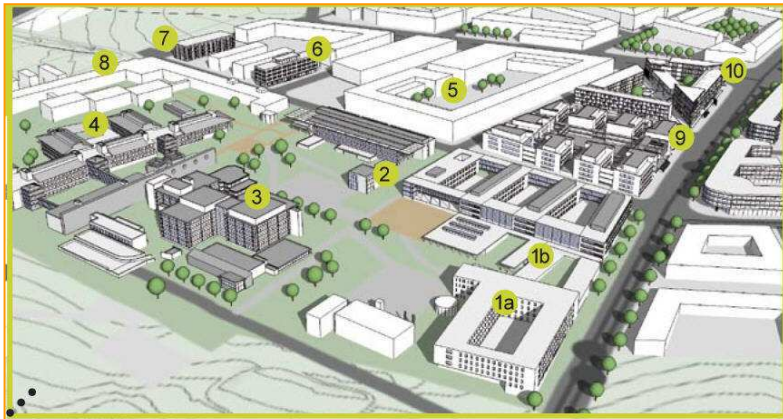- **Contact** with **top decision makers** from **politics, business** and **research**

FIZ Frankfurt Innovation Center Expansion

The **local** and **international network** is a key condition for **growth and expansion.**

Growth

- **Science City Frankfurt-Riedberg:**
  - **266 hectars (28 mio. sq ft)**
  - **N° of residents: 15.000**
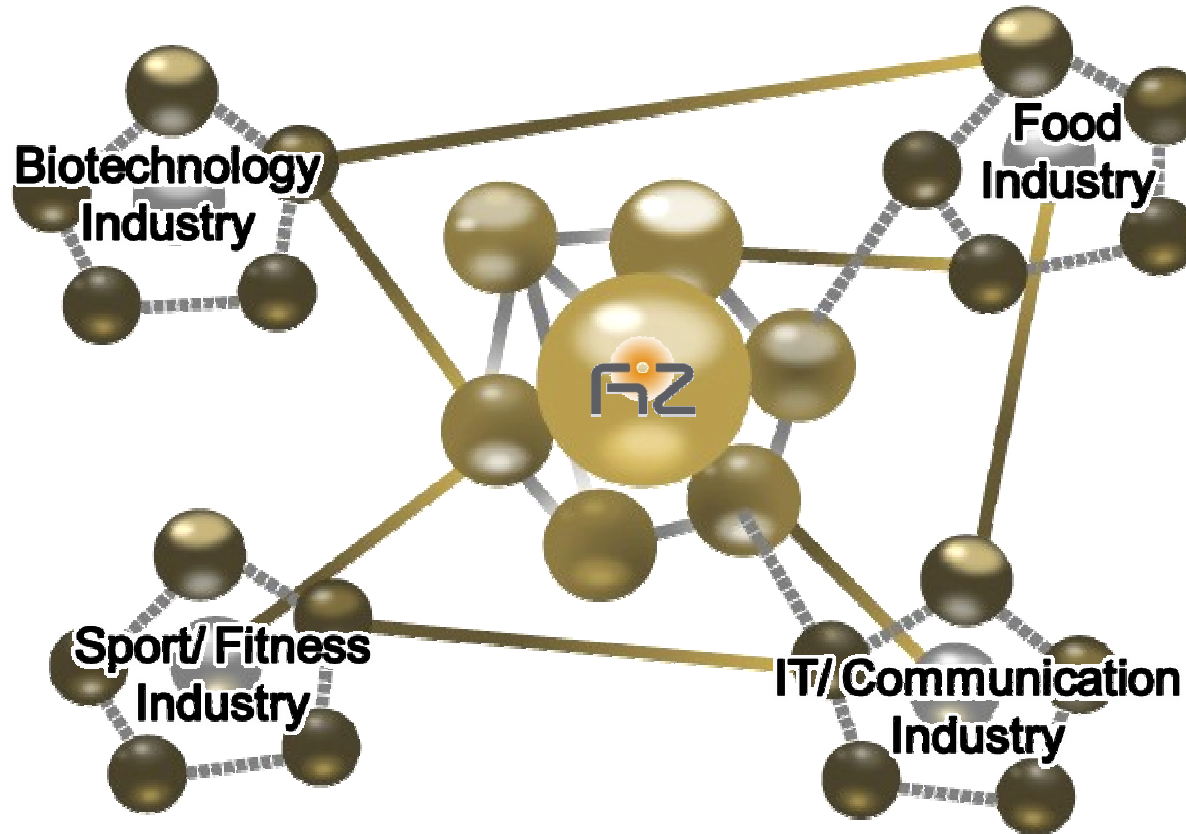 - **N° of working places: 3.000**
  - **N° of students: 8.000**

Internationality

- Acquisition of **national and international small and medium sized companies**

- Continuous **expansion of the network**

FIZ Innovation Landscape for Initializing Cross Innovation

FIZ exploits the opportunities of the innovation landscape as a key driver for **cross innovation** in order to respond to the latest trend of prevention and personalization.

How do we do it?

- We have subject experts who assist our tenant companies in their research and in finding collaborations to commercialize their products.

- They are required to analyze a large number of patents in a very short space of time and of course with very high level of accuracy.

- Earlier the companies were using a fully manual approach which gave them a high level of accuracy, but was very time consuming.

Evolvus' intelligent Information mining tool – Infomine©

- Infomine is a software platform to manage large and complex data mining assignments.

- It uses a hybrid approach wherein machine learning based text-extraction software is coupled with assisted expert annotations by human editors.

- Platform defines the structure and the mechanism in which the data is channelized through appropriate software tools and processes to extract relevant information.

- In our experience, Infomine is giving 100% accuracy with best possible efficiency.

Software Platform Process
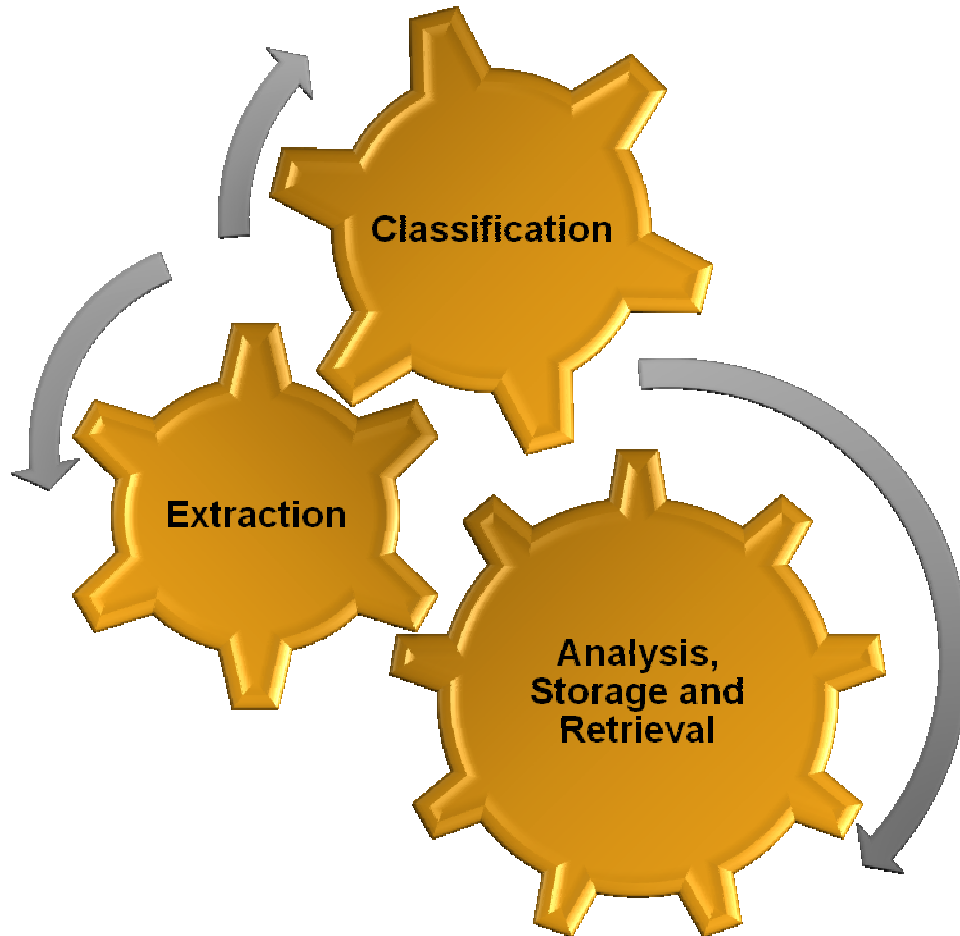
Classification:
Automated process
Based on keywords
Multiple source formats

Extraction:
Segmentation of data
Semi-automated extraction of
relevant content

Analysis, Storage and Retrieval:
Analysis by expert human editors
Near 100% accuracy
Normalized data storage
Easily query and extract information

Patent Data Extraction and Annotation

- **Classification**
  - Web Crawler technology to download abstracts from patent websites
  - Classification of patent documents based on keywords
    - Search based on dictionary of keywords e.g. Markush, Antagonist, Therapeutic, Sidechains, Assay etc
    - This Dictionary is customizable
  - Organization of patent documents based on multiple parameters
    - E.g. based on patent owner, year etc…
  - Conversion of documents to standardized format

- **Extraction**
  - Break documents into segments
  - Use of proprietary, heuristic based learning algorithm
  - Software assisted by team of expert human editors
    - Automation cannot be 100% accurate
    - Editors analyze segments and perform necessary corrections

Patent Data Extraction and Annotation

- **Analysis**
  - Push each segment to team of expert human editors
  - Human editors re-analyze the segment and extract relevant information
  - Different teams with different domain expertise work together
    - Biology, Chemistry, Pharmacology etc…
  - Seamless and simultaneous contribution
    - Two teams can work on the same segment in parallel improving efficiency

- **Storage & Retrieval**
  - All information converted into normalized form for standardization
  - Standardized mechanism to query and retrieve information
  - Ability to create complex queries to retrieve only the data of interest

United States Patent [19]

Patwardhan

US005494668A

[11] **Patent Number:** 5,494,668

[45] **Date of Patent:** Feb. 27, 1996

[54] **METHOD OF TREATING MUSCULOSKELETAL DISEASE AND A NOVEL COMPOSITION THEREFOR**

[76] Inventor: **Bhushan Patwardhan**, 1471 Shukarwar Peth, Bhaumaharaj Bol, Poona 411 002, Ind.

[21] Appl. No.: **273,189**

[22] Filed: **Jul. 11, 1994**

[51] Int. Cl.^{6} A61K 35/78
[52] U.S. Cl. **424/195.1**; 514/825; 514/886; 514/925
[58] **Field of Search** 424/195.1; 514/825, 514/886, 925

[56] **References Cited**

U.S. PATENT DOCUMENTS

4,324,785   4/1982   Stevens 424/195.1
4,906,471   7/1990   Liu 424/195.1
5,064,823   11/1991   Lee et al. 514/198
5,120,538   6/1992   Oie 424/195.1

Primary Examiner—John W. Rollins

[57]   **ABSTRACT**

A method of treating degenerative musculoskeletal diseases such as rheumatoid arthritis and osteoarthritis in an animal, typically a human, comprises administering to the animal, typically enterally, in a convenient dosage form, a therapeutically effective amount of the beneficiated extracts of the plants ASHWAGANDHA (*Withania somnifera*), SALLAI GUGGUL (*Boswellia serrata*), TURMERIC (*Curcuma longa*), and GINGER (*Zingiber officinale*) in a predetermined proportion relative to each other with or without other biologically active inorganic ingredients, such as zinc sulphate. The beneficiated plant extracts are made in accordance with a novel process which is also disclosed.
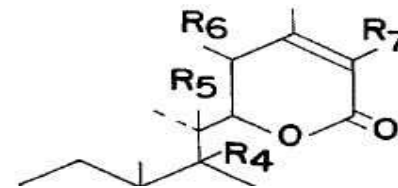
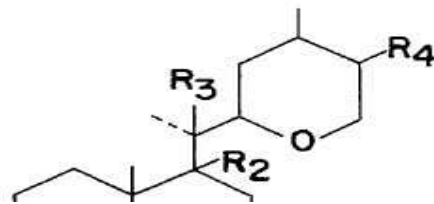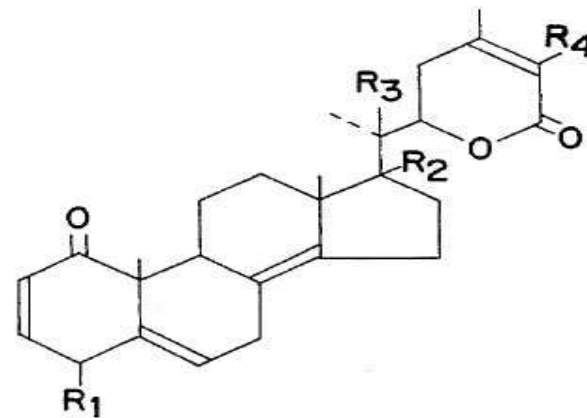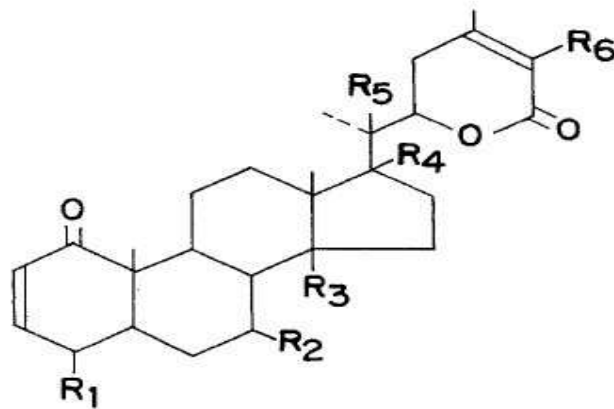**12 Claims, 3 Drawing Sheets**

Sample Patent Document Abstract

U.S. Patent   Feb. 27, 1996   Sheet 1 of 3   5,494,668

Conclusion

- We found Infomine to serve our purpose in annotating very large unstructured patent data.

- FIZ plans to expand this in other scientific areas (Proteomics).

- Application of Infomine has improved our turn around time significantly.

Thank you for your attention!