



**Information
Retrieval
Facility**

IRF Industrial Cooperation Opportunities

ICIC 2010

Marie-Pierre Garnier, Corporate Liaison Manager

mp.garnier@ir-facility.org

www.ir-facility.org

<http://ir-facility.net>

The Information Retrieval Facility (IRF)

- Founded in **2007** in **Vienna, Austria**
- **Independent non-profit** research institute
- Member-based organization with over **250 scientific members from around the globe**
- **Managed by a Scientific Board** composed of leading IR experts
- Chairman of this Board is **Keith van Rijsbergen**, one of the founders of modern Information Retrieval

OBJECTIVES:

- Promotes and facilitates research in **large scale information retrieval**
- Fosters **knowledge transfer** between academia and industry



**Information
Retrieval
Facility**

Founding Members



CSIRO



The
University
Of
Sheffield.

Carnegie Mellon



UNIVERSITY
OF
Massachusetts
Amherst



UNIVERSITY
of
GLASGOW

UNIVERSITÄT
DUISBURG
ESSEN



SINGAPORE
MANAGEMENT
UNIVERSITY

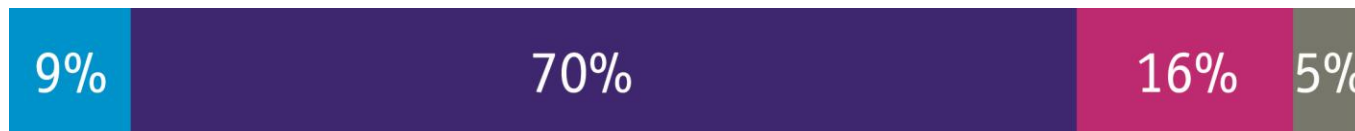
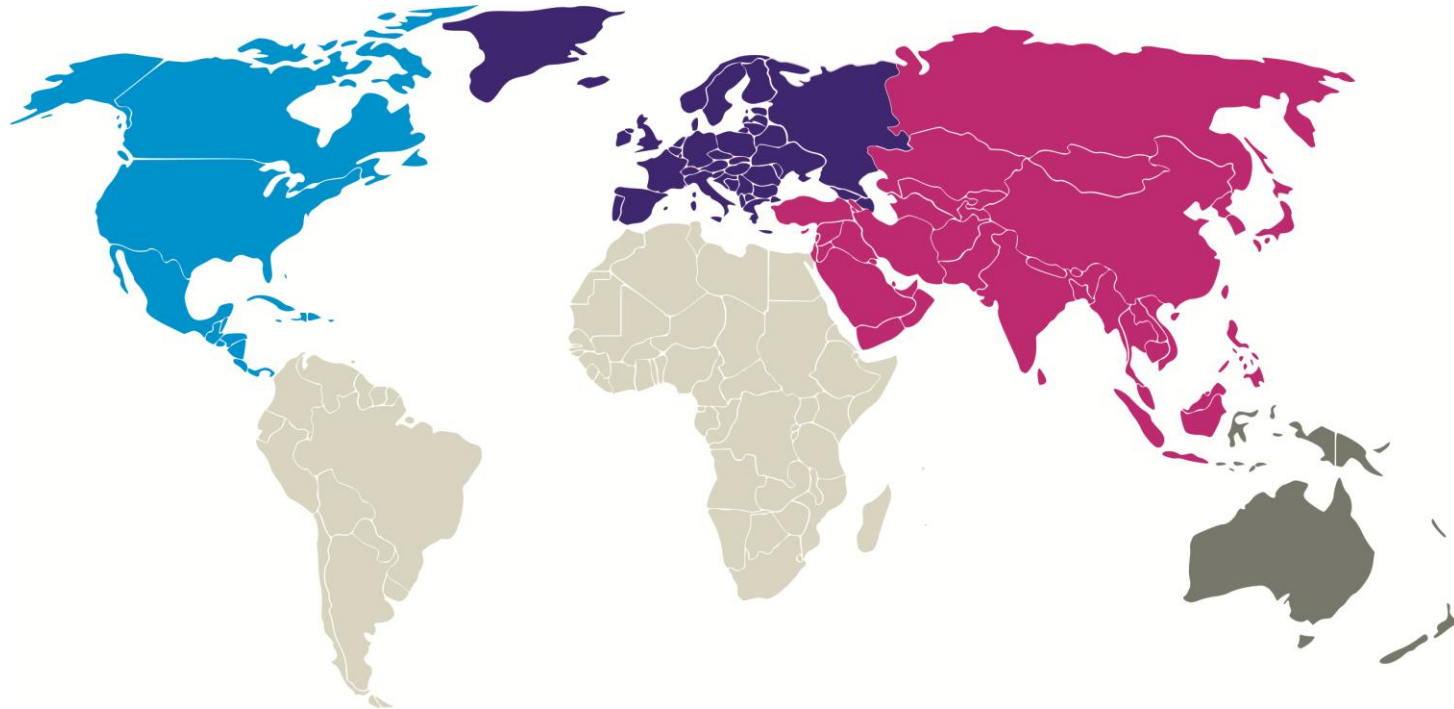


UNIVERSITÉ
JOSEPH FOURIER
SCIENTIS TU ET AVANCE TU MEDICINAM

Geographical Distribution of IRF Members



Information
Retrieval
Facility



What is Information Retrieval (IR)?



**Information
Retrieval
Facility**

IR is the science of searching:

- For documents, for information within documents, and for metadata about documents
- Relational databases and the World Wide Web

IR is inter-disciplinary:

- Computer Science
- Mathematics
- Library Science
- Information Science
- Information Architecture
- Cognitive Psychology
- Linguistics
- Statistics

IR applications include:

- Search Engines
- Digital Libraries
- Information Discovery
- Alert & Route
- Automatic Classification & Annotation

Domain-specific applications of IR include:

Retrieval of:

- Patents
- Legal documents
- Scientific & technical documents
- Images
- Chemical structures
- Market information

IRF Key Research Areas



Information
Retrieval
Facility

Multiple
Indexing

Automated
Text
Annotation

Automated
Information
Extraction

Automated
Document
Categorization

Complex
Professional
Search
Strategies

Image
Retrieval

Interfacing
Professional
Users

Statistical
Machine
Translation

Current Research Projects



**Information
Retrieval
Facility**

PLuTO: Patent Language Translations Online

- Funded by the EU (ICT Policy Support Programme). Duration: 2010-2013
- Consortium Partners: CNGL/Dublin University, ESTeam, Cross Language, WON
- Goal: Develop a collaborative online translation platform for patents

Key Technologies:
Machine Translation,
Translation Memory,
Information Retrieval

KHRESMOI: Medical Information Analysis and Retrieval

- Funded by the EU (ICT FP 7 Programme). Duration: 2010-2014
- Consortium Partners: HES-SO, Atos, Dublin University, Ontotext, University of Sheffield + 7 additional European partners
- Goal: Develop a multimodal, multilingual search and access system for biomedical information and documents

Key Technologies:
Automated Information
Extraction, Automated
Image Analysis and
Indexing,
Linking of Unstructured
Documents with
Structured Knowledge
Base, Cross-Language
Search, Adaptive User
Interfaces

Current Research Projects



**Information
Retrieval
Facility**

Service Detective

- Funded by the Austrian FFG (FIT-IT Programme). Duration: 2009-2011
- Consortium Partners: University of Innsbruck, Seekda
- Goal: Automatic discovery of web services

Key Technologies:
Web Crawling,
Semantic Annotation

LarKC: Large Knowledge Collider

- Funded by the EU (ICT FP 7 Programme). Duration: 2008-2011
- Consortium Partners: STI Innsbruck, AstraZeneca R&D, CEFRIEL, Max Planck Institute + 8 additional European partners
- Goal: Build an integrated platform for large scale semantic computing

Key Technologies:
Large Scale
Information Retrieval,
Semantic Computing

Other Research Activities



**Information
Retrieval
Facility**

- Phrase-based SMT technology (MOSES)
- Chinese to English automatic translation, customized for patents
- 4 million bilingual aligned sentences obtained from human translated patents

**Statistical
Machine
Translation**

Development of a technique to enable efficient image searches for a variety of technical drawings, such as flow charts, block diagrams, time charts and graph plots

**Image
Retrieval**

CLEF-IP (cross-lingual search of patents) and TREC-CHEM (retrieval of chemical documents)

**Evaluation
Tracks**

- Development of a framework for professional search based on the workflow paradigm
- Goal: Translate extensive information search tasks in an easy-to-use graphical interface

**Interfacing
Professional
Users**

Why cooperate with the IRF?



**Information
Retrieval
Facility**

➔ **Scientific Excellence**

Access the know-how of leading IR scientists

➔ **Independence**

Benefit from our neutral guidance and evaluation know-how

➔ **Light Weight Projects**

Get quicker results and lower your development costs through the adaptation of existing academic tools and open source software

➔ **Result Ownership**

Decide for proprietary, shared or open research projects

➔ **Confidentiality**

Early NDA agreements protect your know-how and data

➔ **Costs Savings**

Reduce the project costs by joining the IRF as Project or Strategic Member

IRF Industrial Cooperation Opportunities



Information
Retrieval
Facility

Discovery Workshop

- Identification of the retrieval challenges in your company

IRF Services

- IR Training & Knowledge Transfer
- Expert Finding
- R&D Projects
- Prototyping

IRF Industry Membership

- Access Membership
Use the IRF infrastructure for experimentation and testing
- Project Membership
Buy IRF services at discounted rates
- Strategic Membership
Influence the scientific direction of the IRF and benefit from a research fellowship at the IRF

IRF Services



**Information
Retrieval
Facility**

IR Training

The IRF organizes in-house trainings dedicated to the optimization of search strategies and processes to achieve business relevant search results from the Internet and other information sources.

Expert Finding

The IRF manages a pool of experts in the fields of information retrieval, natural language processing and semantic technologies, all of which have been selected for their proven track record in their area.

We find the experts best suited to your needs!

Information Retrieval

Information Extraction

Machine Translation

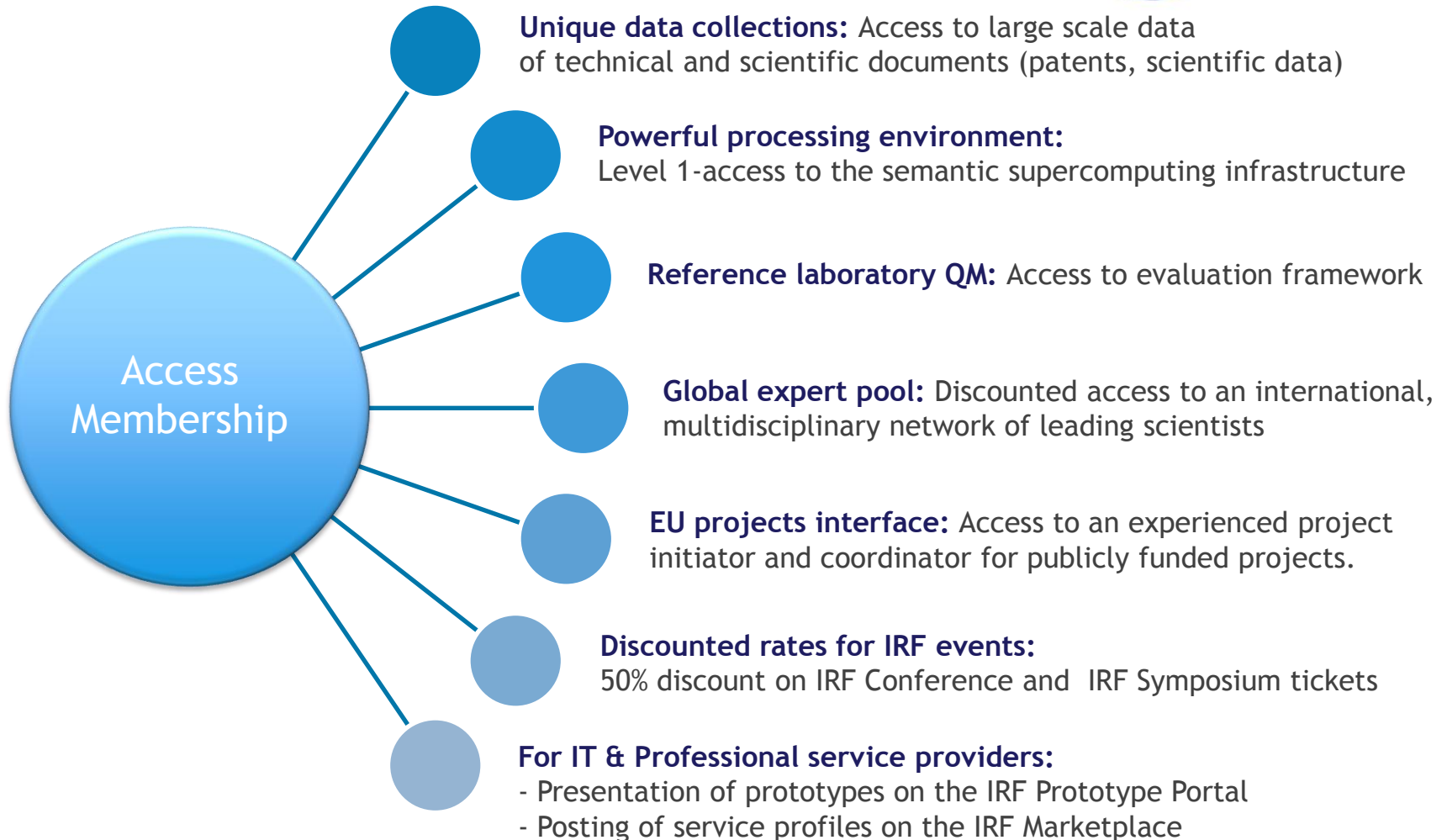
We offer to:

- Conduct and manage research projects
- Write scientific and technical reports
- Build prototypes
- Evaluate tools and systems
- Produce metadata
- Provide coaching for public funding (EU & national)
- Build and coordinate research consortia

Access Membership



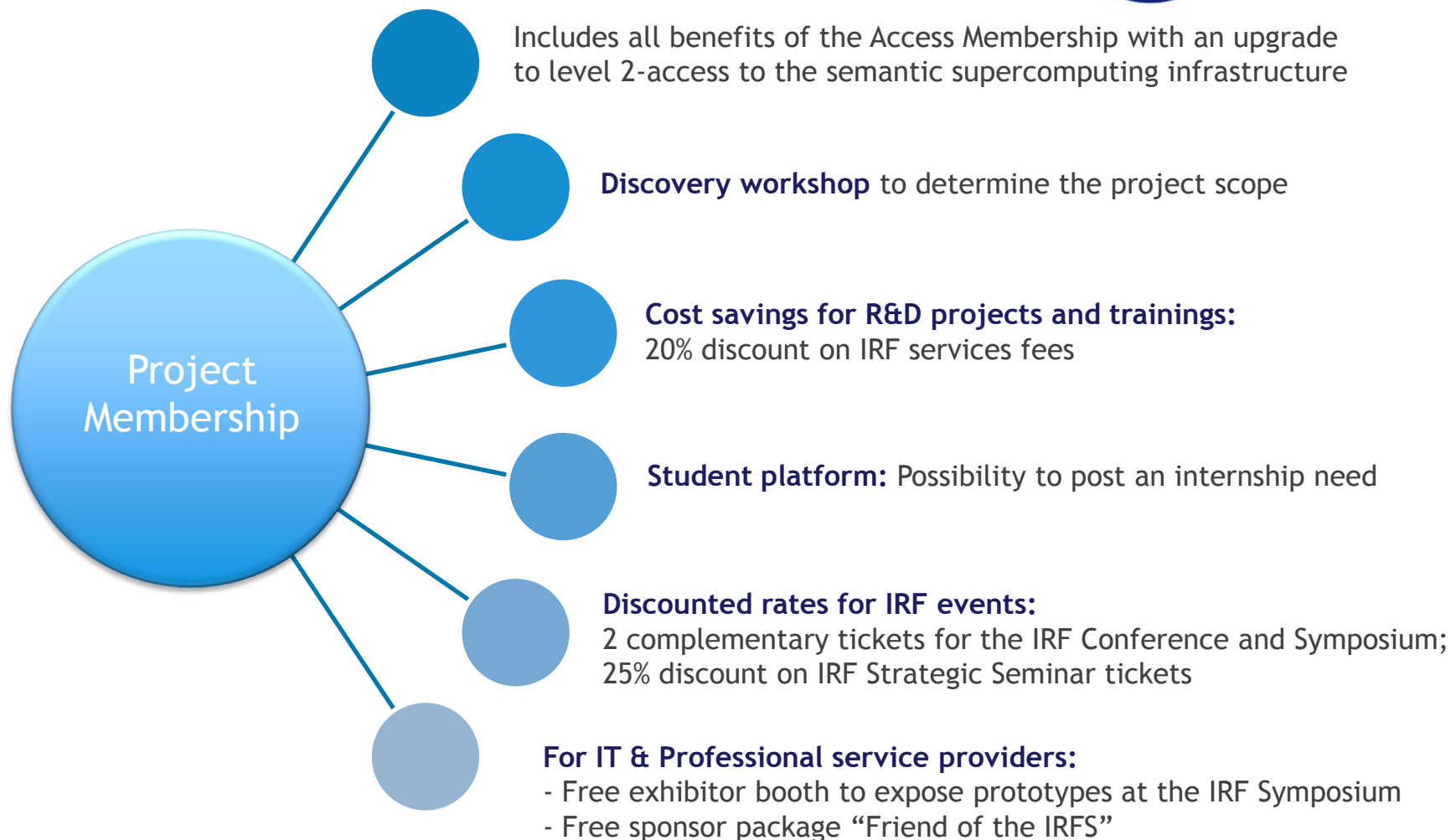
**Information
Retrieval
Facility**



Project Membership



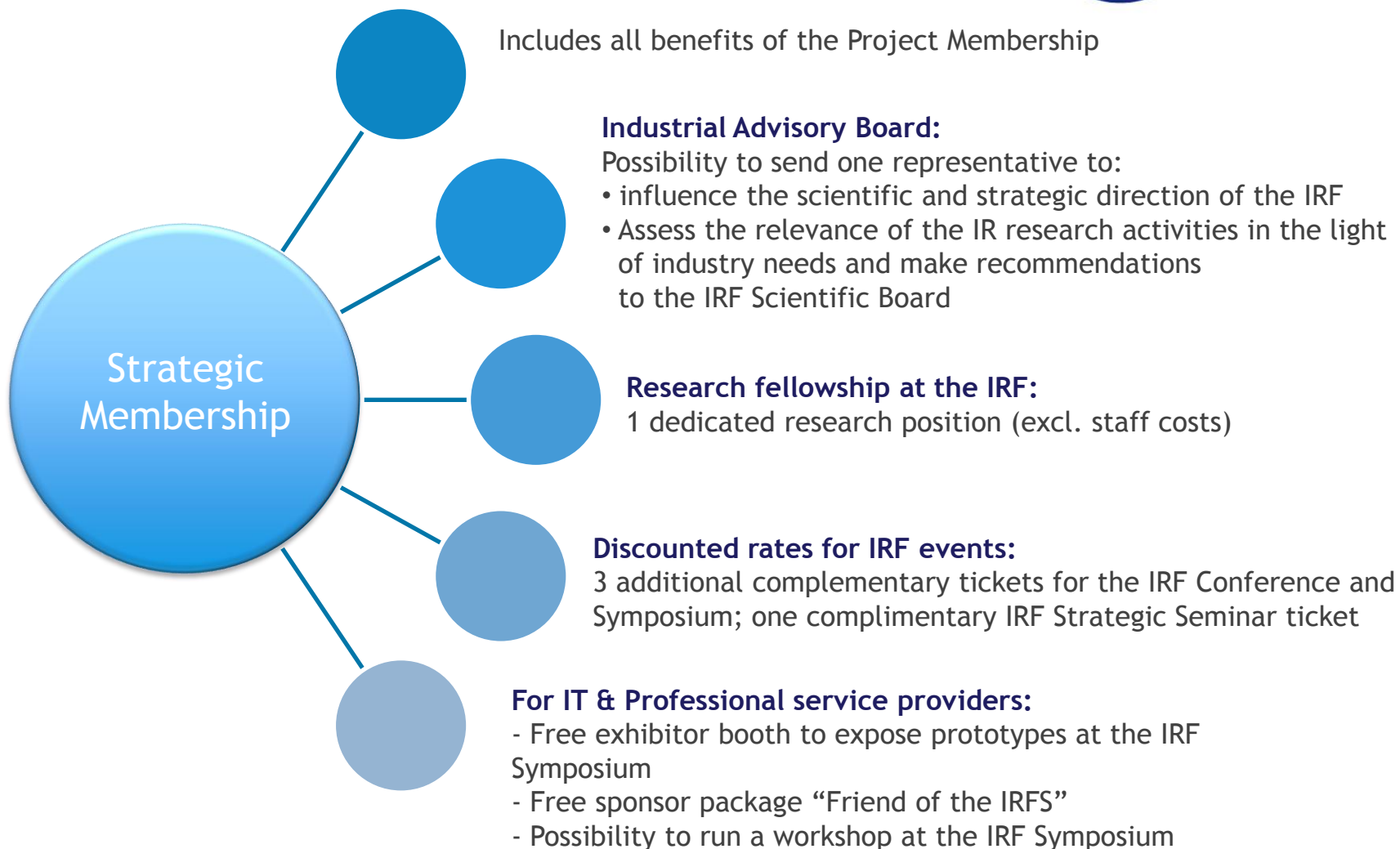
**Information
Retrieval
Facility**



Strategic Membership



**Information
Retrieval
Facility**



Data Collections -Example



**Information
Retrieval
Facility**

MAREC

Consisting of 19 million patent documents in different languages, that are normalized to a highly specific XML format, MAREC is the **first standardized patent data corpus for research purposes**.

MAREC serves as a global storage facility for high quality scientific, technical and business information.

MAREC at-a-glance

- 19 million XML documents
- From 4 patent organizations:
 - European Patent Office (EPO)
 - World Intellectual Property Organization (WIPO)
 - United States Patent and Trademark Office (USPTO)
 - Japan Patent Office (JPO)
- Patent applications and granted patents from 1976 to 2008
- Unified fields, numbering scheme and citation format
- Comparable corpus

Industry Membership provides access and support to the IRF document collections such as MAREC.

LEONARDO

A Framework for Professional Search



**Information
Retrieval
Facility**



leonardo

Leonardo is a premium toolkit framework for retrieval and analysis tasks:

- Designed to enable the elimination of data and technical barriers so that your experts can focus intelligently on the task at hand and share their results with others at all steps of the process.

Leonardo sets a new standard for the human intelligence enterprise:

- It can plug into virtually any data source, enabling you to connect different kinds of information and integrate different tools.

Leonardo has been custom-engineered:

- It is built on a IR Workflow model, which allows retrieval tasks to be solved similarly in a scalable and extensible environment.

Leonardo has an easy-to-use graphical interface :

- It allows to use complicated search processes without IR knowledge.

JUNE 6-9, 2011 | VIENNA | AUSTRIA

4th IRF SYMPOSIUM

Collaboration Experimentation Evaluation



Information
Retrieval
Facility

- 3 plenary sessions: **Image Retrieval, Machine Translation and Interfacing**
- **Panel discussions** at the end of each session
- Presentations involving **demonstrations** and **specific use cases**
- Exhibition showcasing **new technologies** and **prototypes**
- **PatOlympics** - Second edition of the interactive prototype evaluation competition
- **Biomedicine** and **Trademark** tracks for new perspectives on how to improve patent search and analysis
- Second edition of the **IRF Scientific Conference**, on 6 June 2011



www.irfs.at