# MOLECULAR CONNECTIONS PVT LTD.

**KNOWLEDGE SOLUTIONS FOR ACCELERATED RESEARCH**

## Get the most of your content !!!

### Adapting to the new age consumer needs

KNOWLEDGE SOLUTIONS
FOR ACCELERATED RESEARCH

# Agenda

- What we are
- Our core offerings
- Content Enrichment
  - Semantic Tagging
  - Content Outlinking

- **Molecular Connections is a good partner for Semantic Tagging Strategy and Content Enrichment**

- **The company has some good people with experience in wet lab for drug discovery**

- **Harry's conference seems to be getting better with Age or maybe it is the new venue which is having an impact.**

KNOWLEDGE SOLUTIONS
FOR ACCELERATED RESEARCH

# Our services offerings

- Data curation for life-sciences companies
- Customised content development
- Abstracting
- Customised Ontology development, cleanup
- Indexing
  - Semi-automatic
  - Manual

KNOWLEDGE SOLUTIONS
FOR ACCELERATED RESEARCH

# About Molecular Connections – Nutshell

- Literature mining, data curation and database development is our core business

- A reputed team with more than 500 man years-experience & expertise in:
  - literature mining, custom content creation, database dev., analytics
  - target validation, creation of target/drug/disease based knowledgebases

- Well-defined knowledge management platforms and work flow systems. ISO certified.

- Scientific & technical team - comprises of experts in - Life Sciences | Chemistry | Bioinformatics | Software development & Engineering

- Few Awards:
  - ✓ "Most promising Enterprise & Best SME of the Year" – 2008 – ICIC/CNBC award winner amongst 300,000 entries
  - ✓ Winner of Deloitte Tech Fast 50 & 500 – Ranked 8th in India and 61st in Asia Pacific
  - ✓ "Business Excellence Award" by IED, Govt. of INDIA

KNOWLEDGE SOLUTIONS FOR ACCELERATED RESEARCH

# Awards & Accolades

- Ranked **8th fastest growing company in India** & 64th fastest in Asia **(Deloitte, Tech Fast 50 & 500)**

- Awarded '**Most Promising Enterprise of the Year' & 'Best SME' from India  - by ICICI- CNBC** ( 325,000 applications)

- First informatics company to go in for INFORMATION SECURITY STANDARD ISO-IEC 27001/2005

- IES conferred "BUSINESS EXCELLENCE AWARD" to Molecular Connections & "UDYOG RATTAN AWARD" to our CEO in 2009
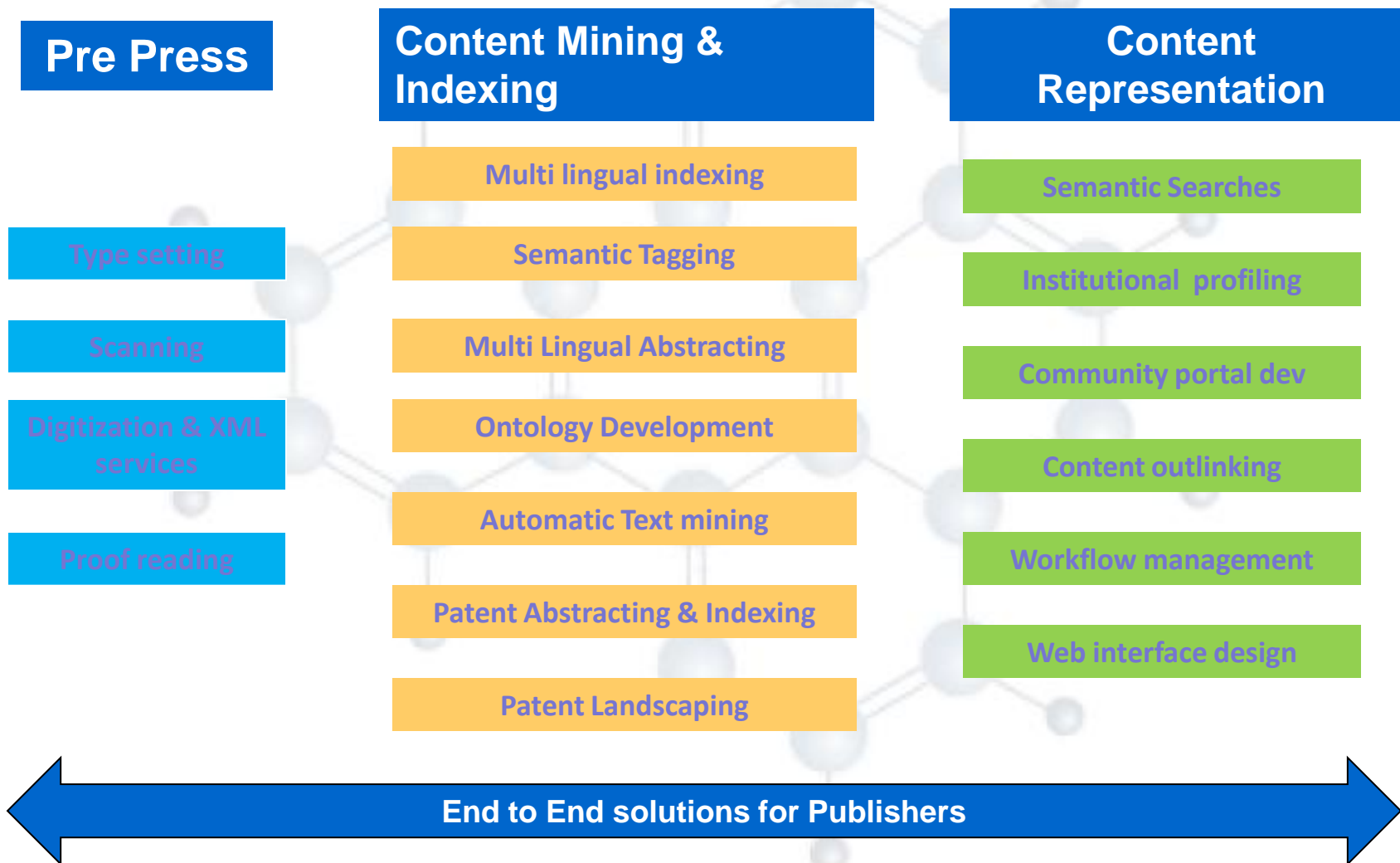
# Select Clientele
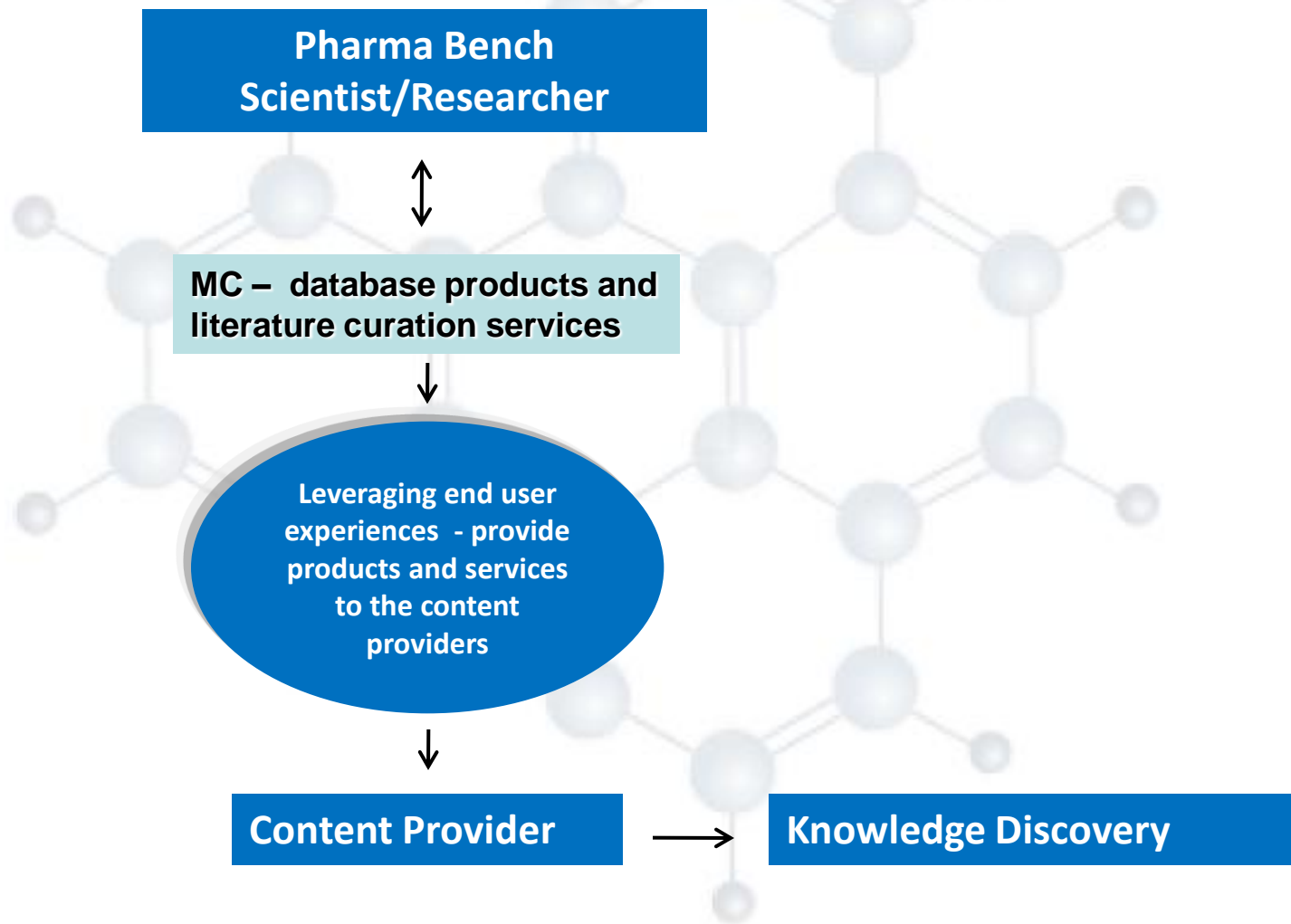
# Curation work examples

- Epigenetics knowledgebase
- Protease target knowledgebase
- Arabdopsis Knowledgebase
- Drug Target Reports
- Drug Toxicity database
- Biomarker database
- Pathway centric database
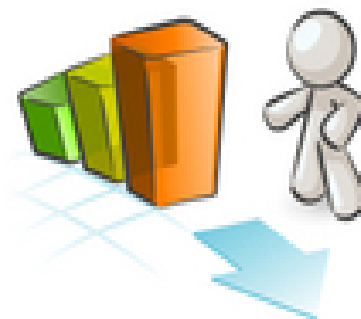
8

# MC- Publishing Solutions

## Pre Press

- Type setting
- Scanning
- Digitization & XML services
- Proof reading

## Content Mining & Indexing

- Multi lingual indexing
- Semantic Tagging
- Multi Lingual Abstracting
- Ontology Development
- Automatic Text mining
- Patent Abstracting & Indexing
- Patent Landscaping

## Content Representation

- Semantic Searches
- Institutional profiling
- Community portal dev
- Content outlinking
- Workflow management
- Web interface design

**End to End solutions for Publishers**

KNOWLEDGE SOLUTIONS FOR ACCELERATED RESEARCH

# Bridging the discovery gap between Pharmas & Content providers



**Pharma Bench Scientist/Researcher**

↕

**MC – database products and literature curation services**

↓

Leveraging end user experiences - provide products and services to the content providers

↓

**Content Provider** → **Knowledge Discovery**

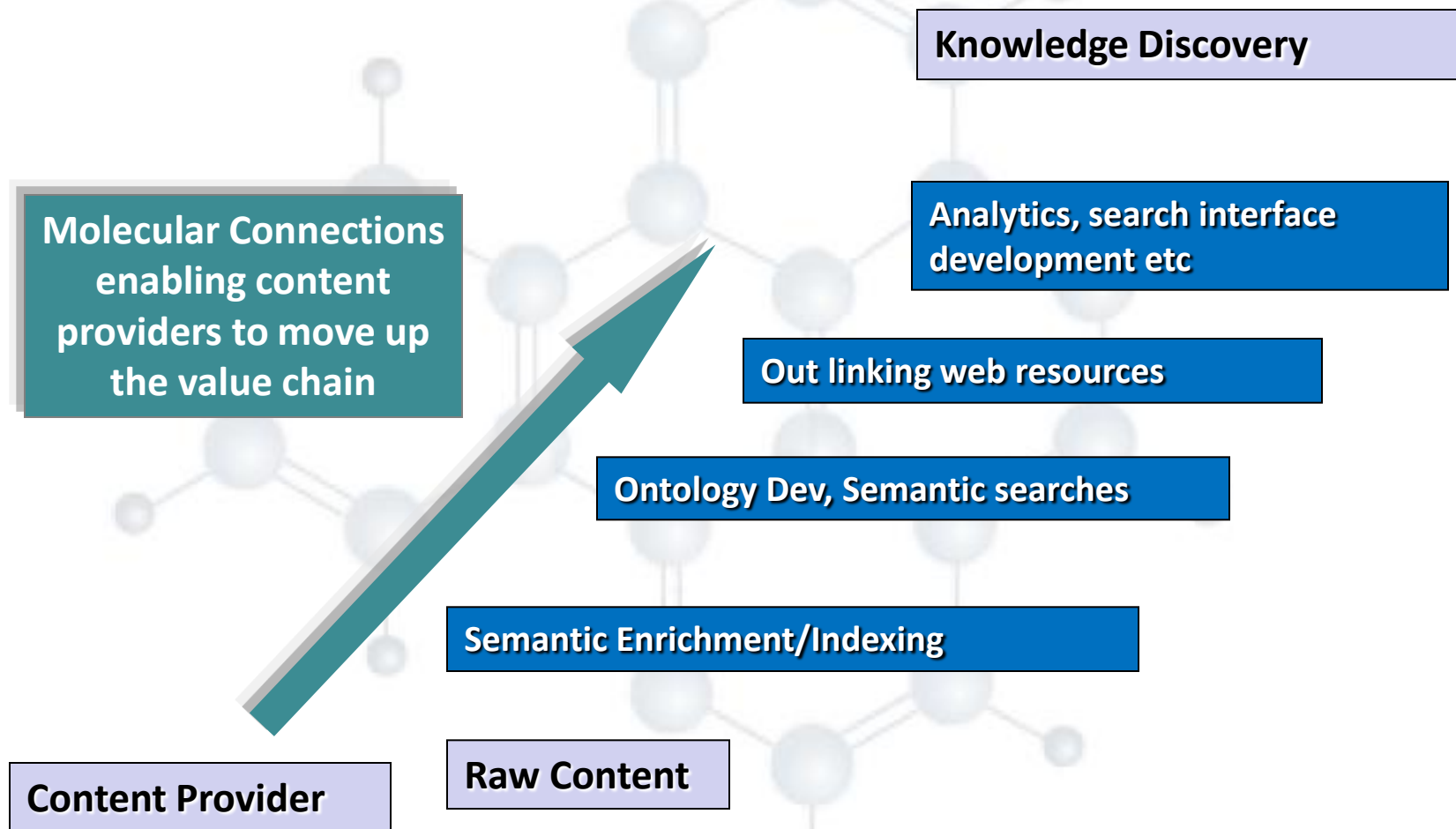KNOWLEDGE SOLUTIONS
FOR ACCELERATED RESEARCH

# Consumer Driven Publishing

## The New age Publishing Consumer needs

– Quicker means to analyze published literature

– Values validations from community & external sources

– Interactions with Authors, locating peers

– Better control over information- enabling business analytics

– Social networking, Collaborations & data sharing

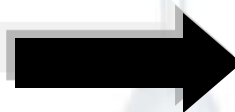– Discover knowledge from the published content
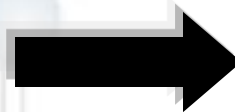
# How can we add value?

**Knowledge Discovery**

**Molecular Connections enabling content providers to move up the value chain**

**Analytics, search interface development etc**

**Out linking web resources**

**Ontology Dev, Semantic searches**

**Semantic Enrichment/Indexing**

**Raw Content**

**Content Provider**

KNOWLEDGE SOLUTIONS
FOR ACCELERATED RESEARCH

# Research/Products by Molecular Connections

| | |
|---|---|
| *Can we obtain, semantically enriched Meta Data for the Published content* | **Semantic Enrichment** |
| *Can I validate the published findings with information from Wikipedia, blogs etc* | **Connect Meta Data with Web resources** |
| *Can I get aggregated content from 100 web sites of my choice everyday* | **MC-Out link- content aggregation from multiple web sites** |
| *Can I make quick hypothesis from the published scientific facts* | **Semantic and Ontology Based Searches** |

KNOWLEDGE SOLUTIONS FOR ACCELERATED RESEARCH

# Innovations in Information management

Semantic Enrichment Meta Data

Ontology Development & Meta Data Mapped to Community accepted Ontologies
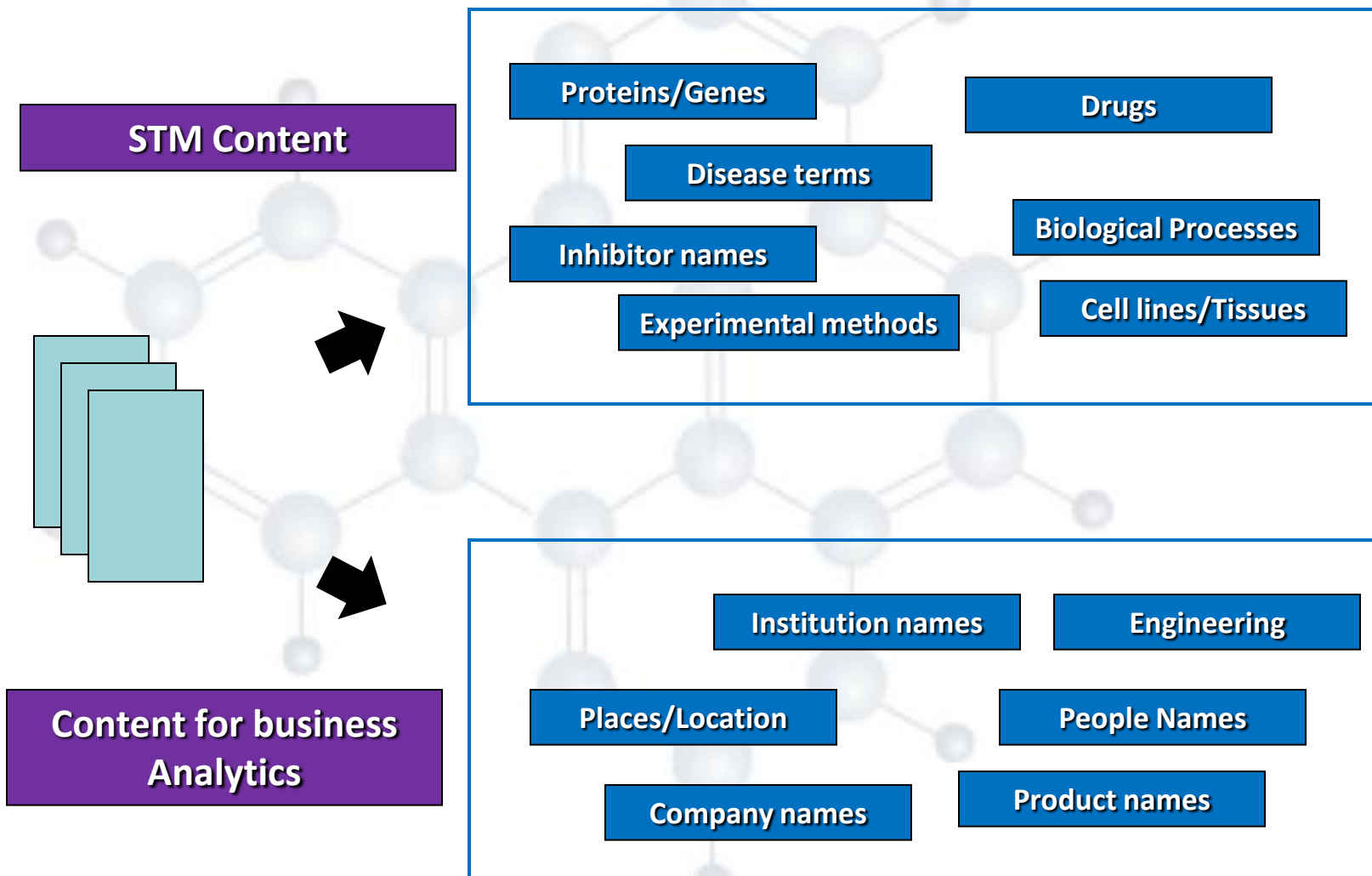
MC-Outlink

Out link the enriched content to aggregated resource of patents, news, wikipedia etc

# Semantic tagging

Involves - content extraction plus mapping it to an ontology/thesaurus

- The ability to identify the relevant entity or concept- either using text mining engines or manually or with a hybrid approach
- To link this information to multiple Ontology's/thesauri/authority files
- Provides commonly accepted set of standards for ease of search and navigation

# MC -Scope of tagging



STM Content

- Proteins/Genes
- Drugs
- Disease terms
- Biological Processes
- Inhibitor names
- Experimental methods
- Cell lines/Tissues

Content for business Analytics

- Institution names
- Engineering
- Places/Location
- People Names
- Company names
- Product names
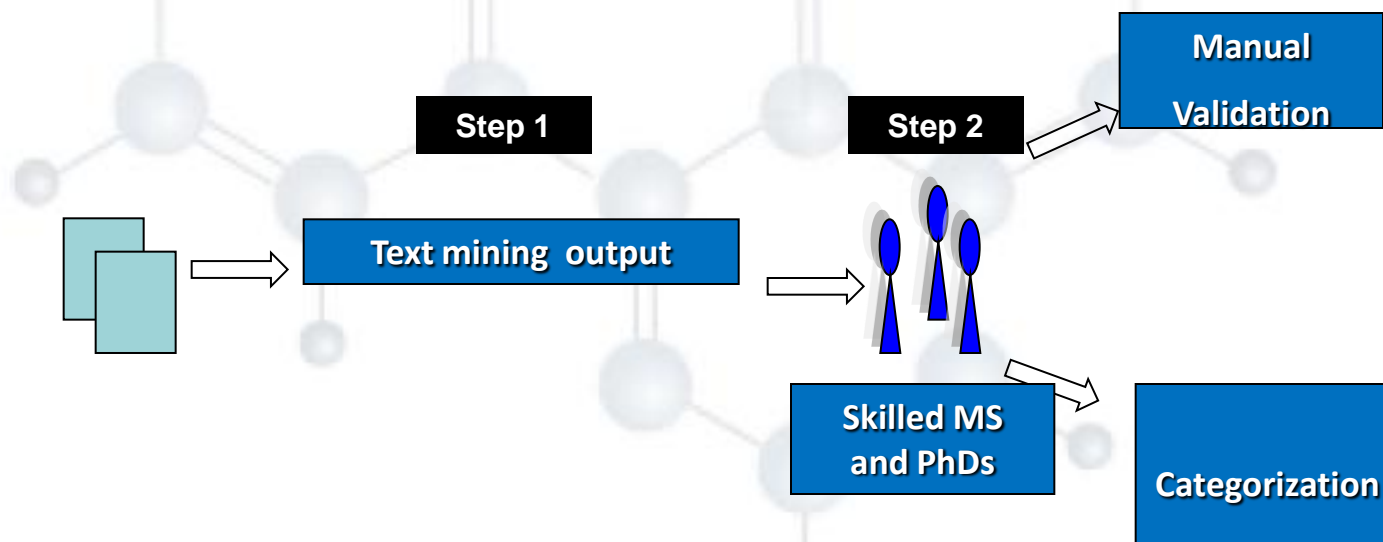
KNOWLEDGE SOLUTIONS FOR ACCELERATED RESEARCH

# Hybrid Approach

In our hybrid approach we have reduced the time involved in the manual annotation effort- by significant use of a mix of text mining followed by manual validation.  Also since we use manual validation over text mined output – it limits the false positives to a large extent.

So that in almost real-time basis we are able to serve you with the accurate manually annotated scientific facts.
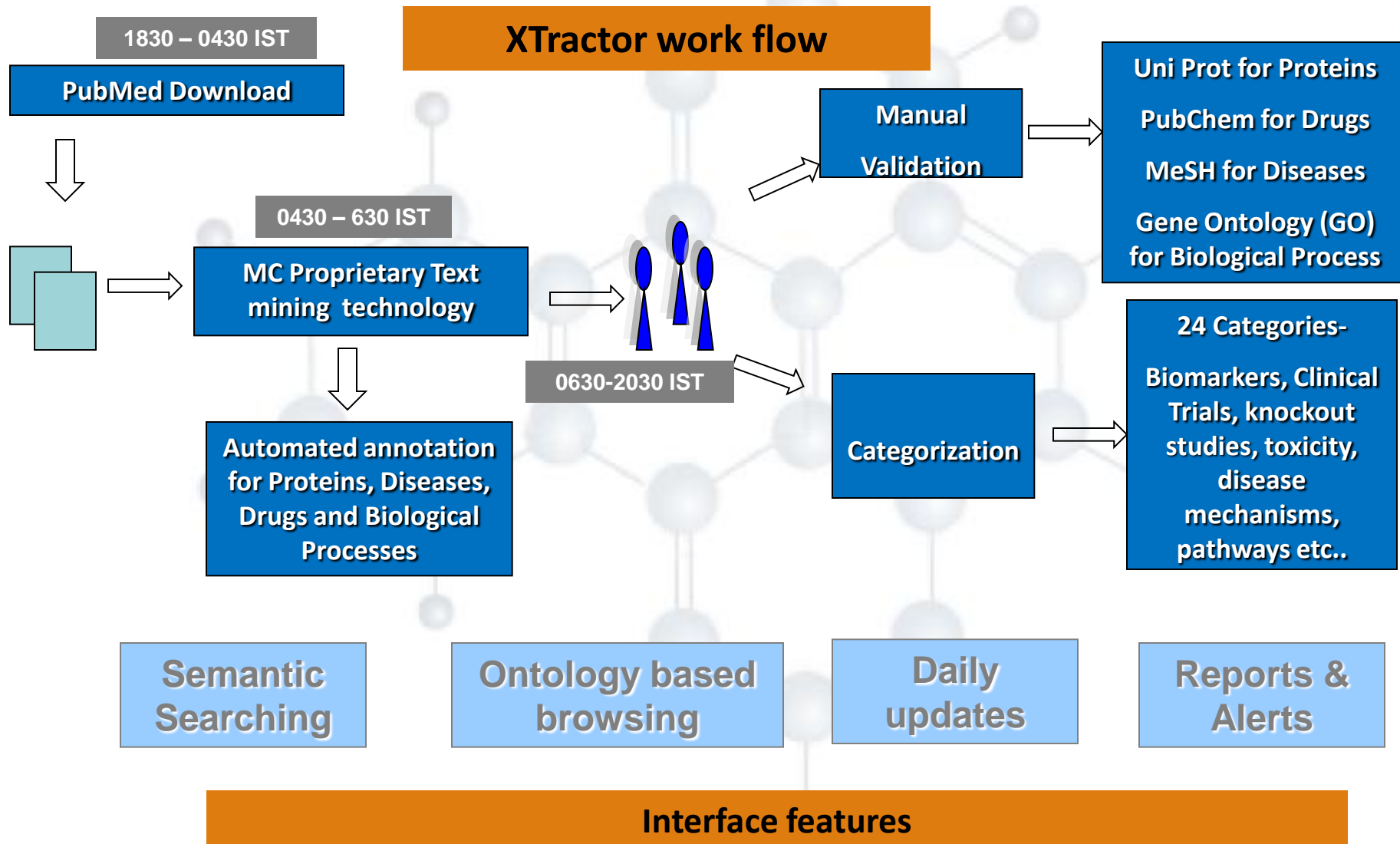
**Step 1**

**Step 2**

**Manual Validation**

**Text mining output**

**Skilled MS and PhDs**

**Categorization**

KNOWLEDGE SOLUTIONS
FOR ACCELERATED RESEARCH

# Case Study

- **Balancing content accuracy & turnaround time**
  - Hybrid approach of Indexing PubMed abstracts
  - Indexing and tagging cycle for each day's PubMed downloads
  - More than 3000 abstracts processed everyday
  - Mapping the metadata to 4 different ontology's
  - Categorization of abstracts to 24 different categorires
  - Highly accurate content

**Each day's PubMed output processed with very high accuracy in a turnaround time of 24hours**

# Work flow & Features

**XTractor work flow**

**1830 – 0430 IST**

**PubMed Download**

**0430 – 630 IST**

**MC Proprietary Text mining technology**

**Automated annotation for Proteins, Diseases, Drugs and Biological Processes**

**0630-2030 IST**

**Manual Validation**

**Uni Prot for Proteins**

**PubChem for Drugs**

**MeSH for Diseases**

**Gene Ontology (GO) for Biological Process**

**Categorization**

**24 Categories-**

**Biomarkers, Clinical Trials, knockout studies, toxicity, disease mechanisms, pathways etc..**

**Semantic Searching**

**Ontology based browsing**

**Daily updates**

**Reports & Alerts**

**Interface features**

19

KNOWLEDGE SOLUTIONS
FOR ACCELERATED RESEARCH

# Ontology expertise (life sciences & chemistry)

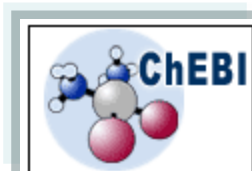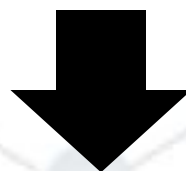**Tagging Meta Data to Ontology/Thesauri enables**

**Common search standards**

| Drugs | Experimental Methods | Biological Process | Diseases | Proteins |
|---|---|---|---|---|

ChEBI

the Gene Ontology

MeSH

UniProt

The IUPAC International Chemical Identifier (InChI™)

PubChem

IUPAC GOLD BOOK

ICD 9

**Publisher's Proprietary Ontology/Thesauri**

**SNOMED CT**

Unified Medical Language System® (UMLS®)

KNOWLEDGE SOLUTIONS FOR ACCELERATED RESEARCH

# Innovations in Information management

**Semantic enrichment of Metadata**

**Ontology Development**

**Meta Data Mapped to Community accepted Ontologies**

**MC-Outlink**

**Out link the enriched content to aggregated resource of patents, news, wikipedia etc**

KNOWLEDGE SOLUTIONS
FOR ACCELERATED RESEARCH

# MC-Outlink

**MC-Outlink enables the end user to obtain highly informative reports from dispersed sources such as web portals, news sites or databases on a single platform**

# Problems

Information Overload

Scattered Information Resources

Not up-to-date with the latest information

Limited Search ability of Databases

Solution

MC-OUTLINK

KNOWLEDGE SOLUTIONS
FOR ACCELERATED RESEARCH

# Core Features

☞ **Integrated access to multiple public database sources**

☞ **Integration of in-house and proprietary resources**

☞ **Keeps you up-to-date with the latest happenings**

☞ **Cuts additional user time in performing web searches**

☞ **Multiple sources enables better decision making**

☞ **Dynamic Search capability**

☞ **Increase ways to query the knowledgebase**

☞ **Ability to generate customized reports**

# Report Structure

# Add value to your content



We could enable you to get comprehensive out linked reports from 200+ web sites for semantically enriched meta data entities

If you are a content provider, Imagine your Literature Databases here

Enable your consumer to get more value added information from web resources

# Value Added Benefits – MC Outlink

☞° Increase web traffic to your content- as pages are better indexed by search engines

☞° Increase individual document sales of content and advertising revenues

☞° Provide your site users with a better search experience

☞° Increase the "stickiness" and page views on your site

☞° Enable discovery of new uses and value within your content

☞° One stop solution for all the related content

☞° Facilitate the discovery of new articles and insights

☞° Easy to implement and maintain

☞° Low maintenance fee

KNOWLEDGE SOLUTIONS
FOR ACCELERATED RESEARCH

# Thank you

## For more information contact us at:

Bangalore : jignesh@molecularconnections.com

London : bill.town@kilmorie.com

New York : john.Reggazi@liu.edu

KNOWLEDGE SOLUTIONS
FOR ACCELERATED RESEARCH