



Organizing Information to support Knowledge Discovery

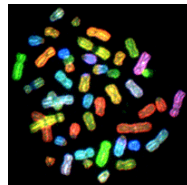
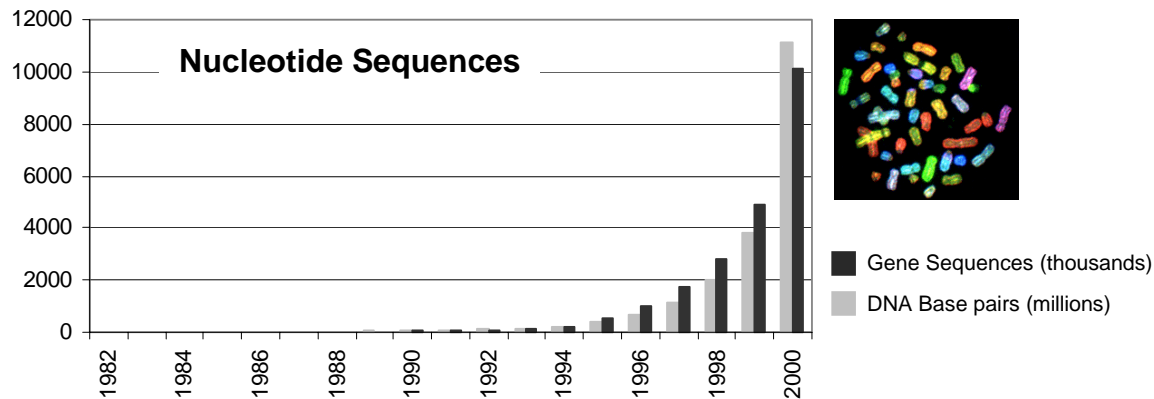
**Edy S. Liongosari
Enterprise Content Management
September 2001**

Introduction

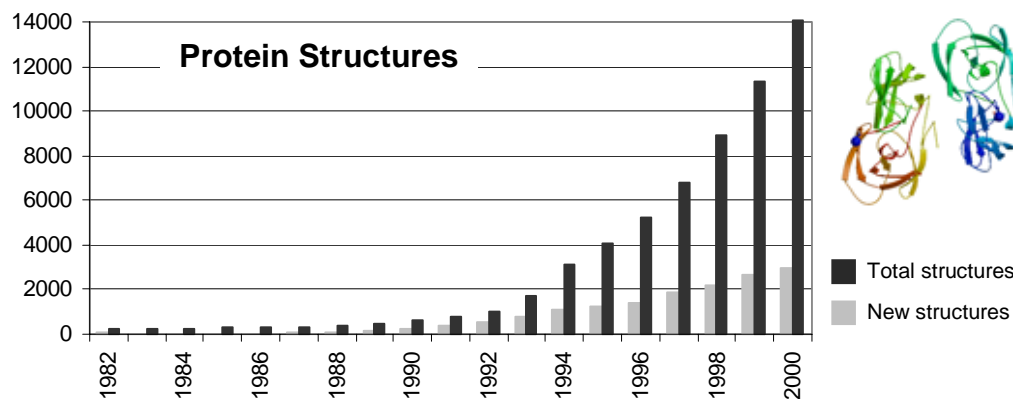
Knowledge Discovery Tool

The volume of genomic data for various organisms in the biological sciences is growing exponentially, resembling a Moore's Law pattern.

Volume



■ Gene Sequences (thousands)
 ■ DNA Base pairs (millions)



■ Total structures
 ■ New structures

Year

■ Model Systems

H. influenzae
S. Cerevisiae
C. Elegans
A. Thaliana
D. Melanogaster

■ Complex Organisms

H. sapiens
M. musculus
R. Norvegicus

■ Infectious Organisms

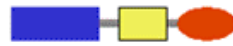
■ Agricultural
Rice

The fundamental data types and analytical methods that scientists are using is becoming increasingly complex and often require multiple visualization applications.

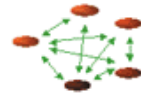
Complexity

MARTKQTARK
STGGKAPRKQ
LATKAARKSA

Sequences



Domain-cartoons
(secondary structure
cartoons)



Similarity groups
or neighborhoods

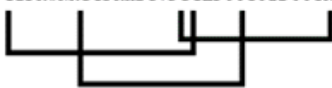
```
GPK-MDGVFCEPY
GGQNSGPTCLASG
SPTSYN---CGR--
SRLHY---DQCT--
QIPYYL---DQCEPL
```

Multiple alignments

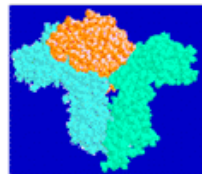


Evolutionary trees

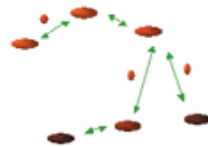
CIPKWNRCGPKMDGVPCCEPYTCTSDYYGNCS



Extended sequences
(e.g. disulphide-topologies)



3D structures



Metabolic pathways



Subunit structures,
ligands



Genomes



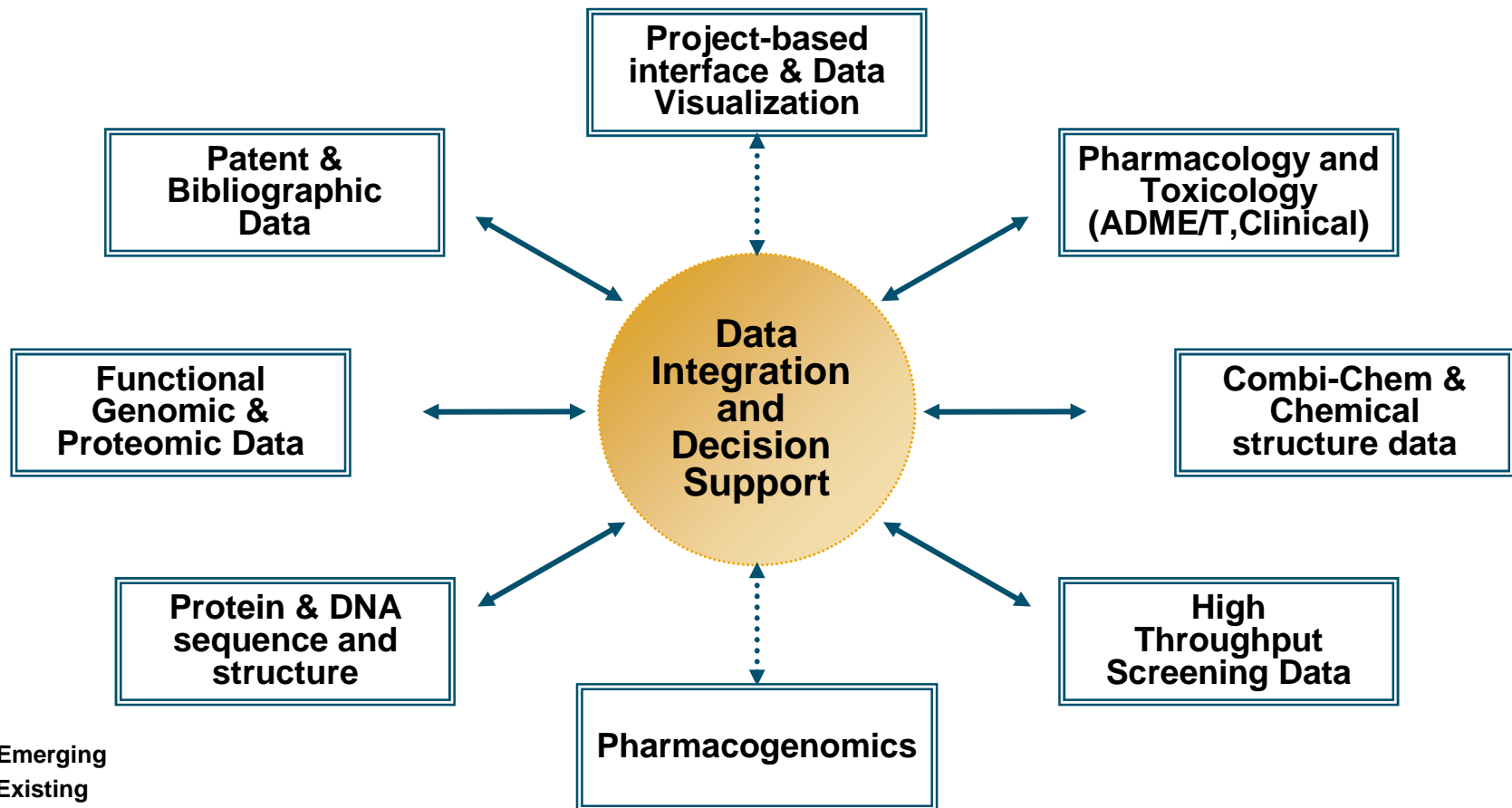
Dynamic
assemblies

Data types

Analysis Methods

There is a tremendous need within R&D institutions to establish a platform for creating and managing the interdependencies between evolving genomics datasets.

Integration Areas



The limitations of current data platform solutions include static searching capabilities, slow translation from data into knowledge, and lack of end-to-end integration.

Current Limitations

- Very “data-centric”, not “information-centric”
- Integration achieved across few datasets
- Results contain no cross references to related items
- Emerging data sources incorporated one at a time



User is limited to the narrow scope of the initial search without further insight into related information

Emerging Data Sources

3 x 10⁹ DNA bases



1 person

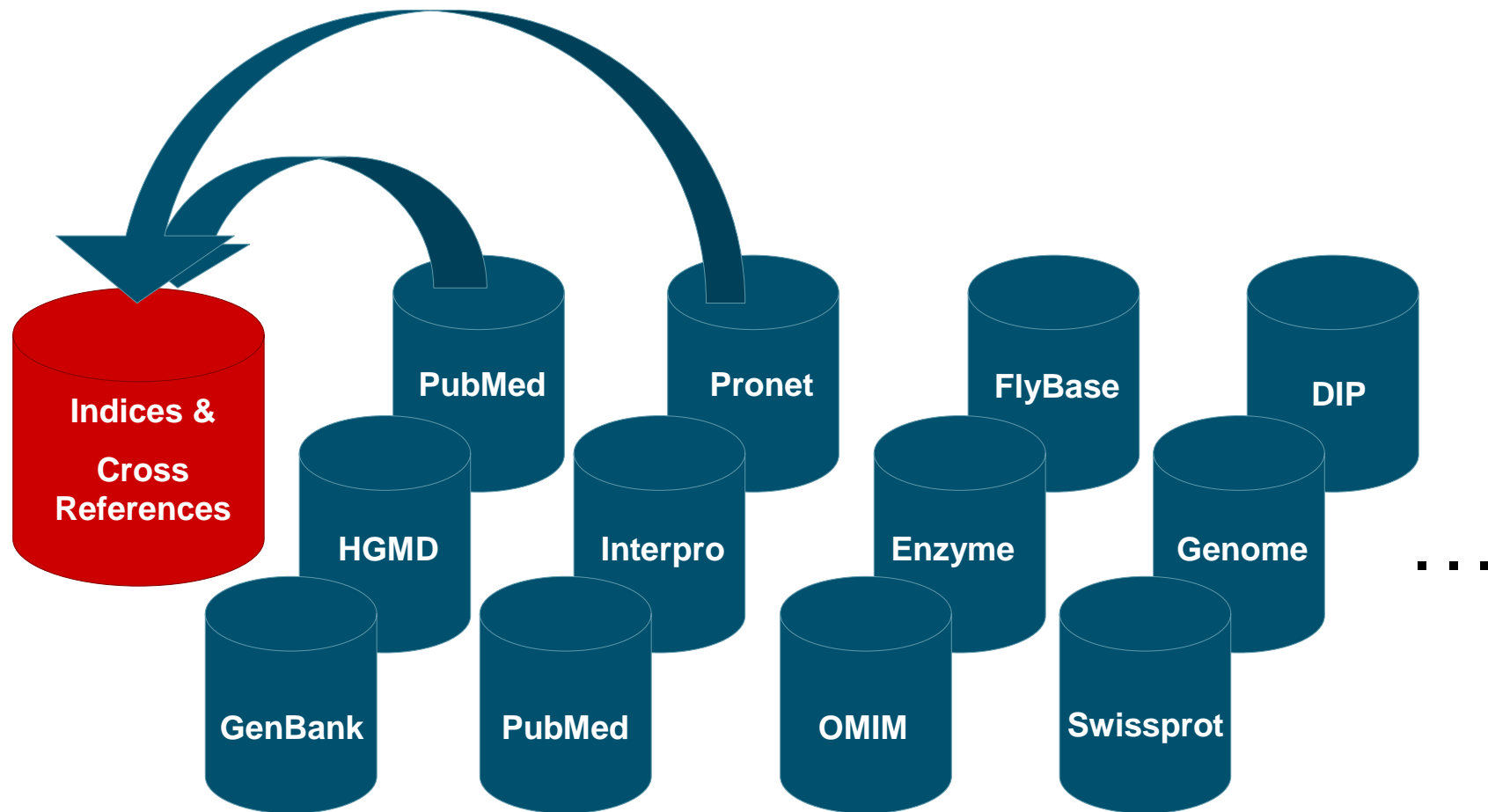
- | | | |
|--------------------------------|-----------------------------|----------------------------|
| ■ Shotgun sequencing | ■ Combinatorial chemistry | ■ SNP analysis |
| ■ Genome Assembly | ■ Single cell HTS | ■ Pharmacogenomics |
| ■ HT Protein synthesis | ■ Immunological reactivity | ■ Pathway predictive tools |
| ■ Tandem MS | ■ Fluorescence technologies | ■ HT pharmacokinetics |
| ■ DNA chips | ■ <i>in silico</i> HTS | ■ Clinical diagnostics |
| ■ HT proteomics | ■ Protein structure | ■ Clinical outcomes |
| ■ Miniaturization technologies | ■ Chemo-informatics | ■ Metabonomics |



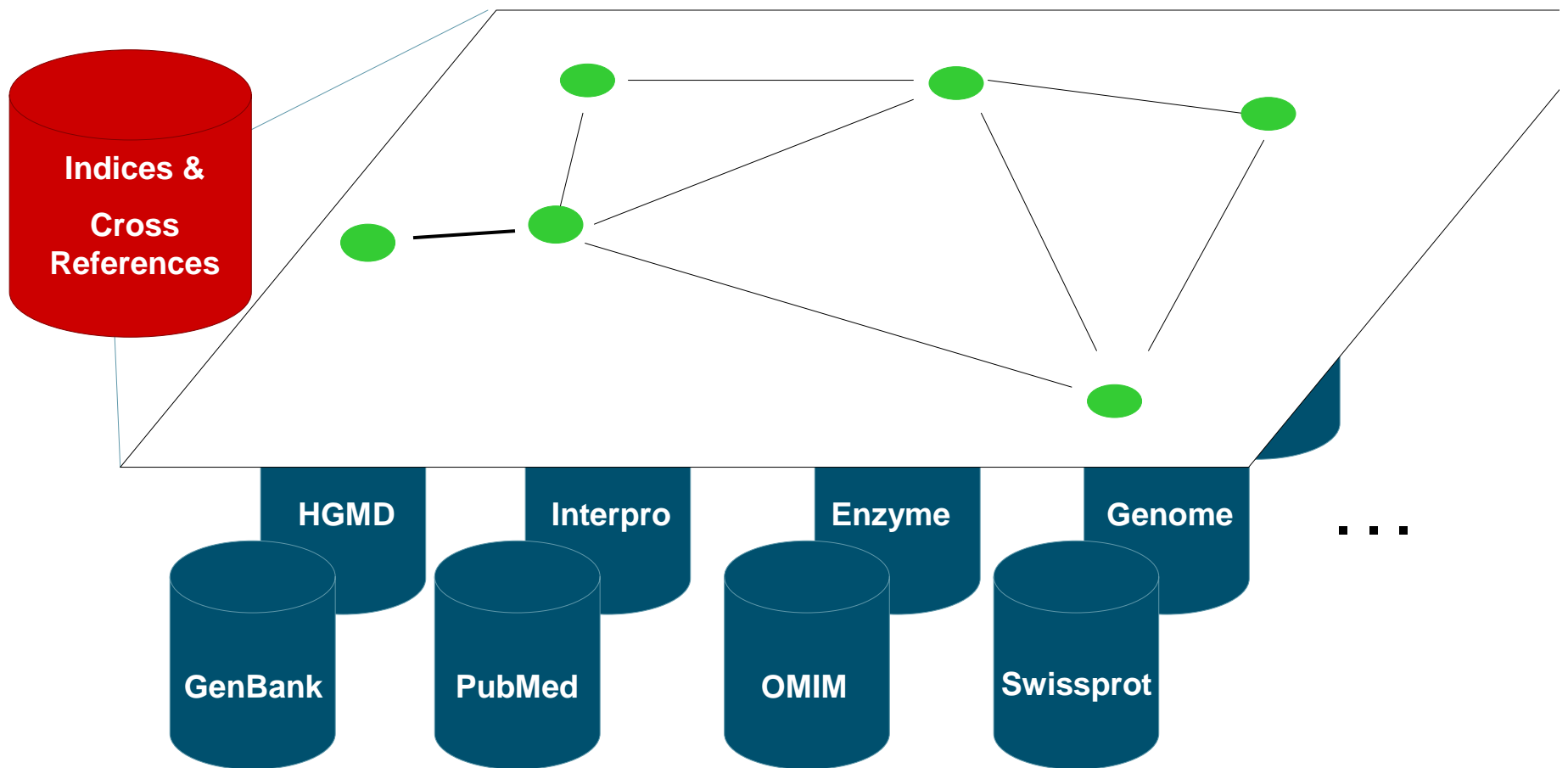
Introduction

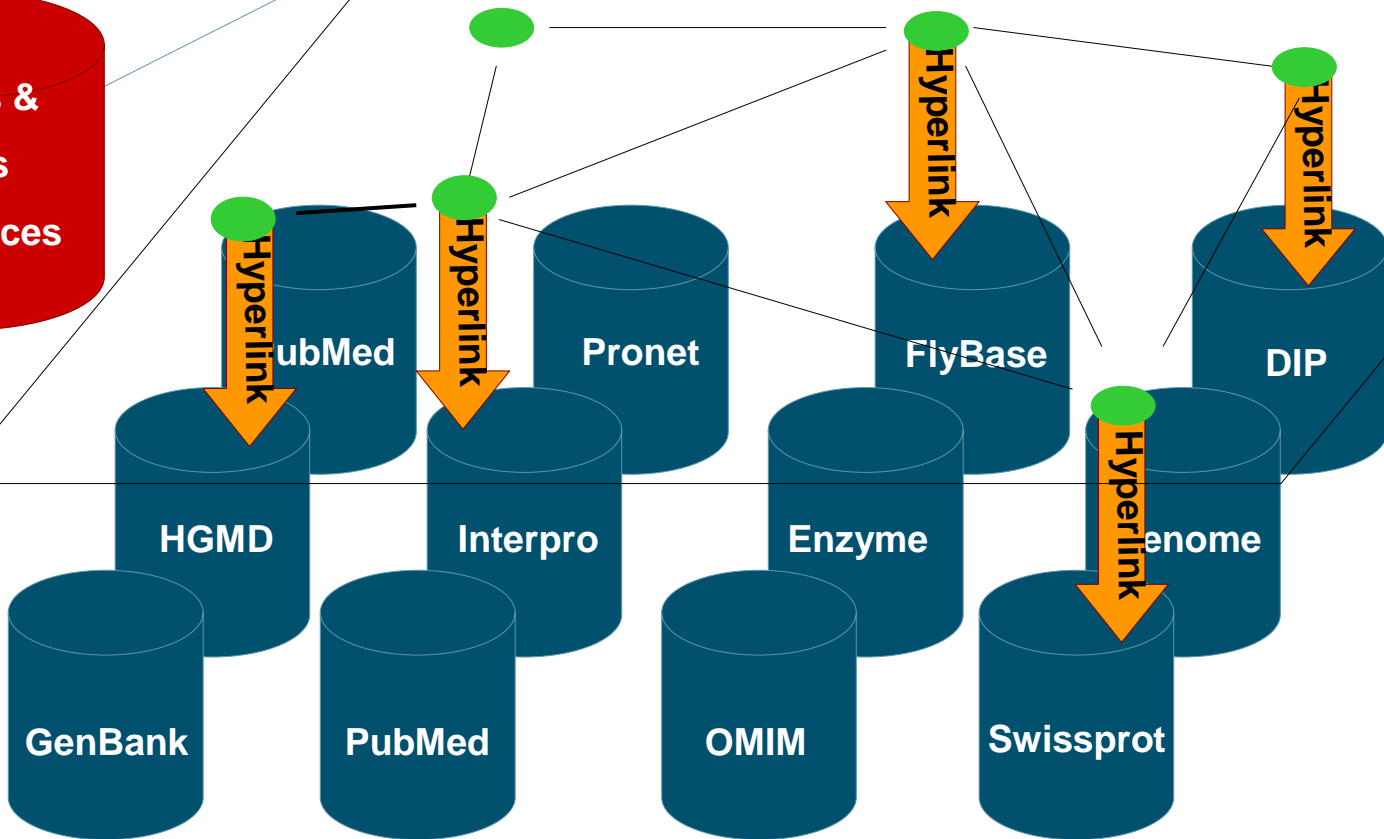
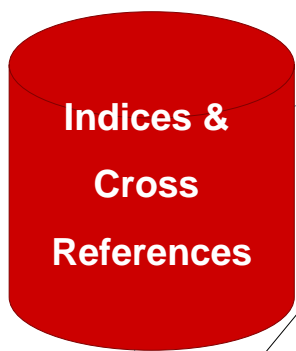
Knowledge Discovery Tool

The Knowledge Discovery Tool indexes disparate internal and external sources such as these biological databases



It indexes the information based on a predefined knowledge model to allow user to quickly navigate and visualize how







Knowledge Discovery

Edy Liongosari
Tuesday, July 24, 2001

Home

Navigation

Monitored
Items

Literature

1 thru 10 of 32

- Intracellular Ca²⁺ dependence of nitric oxi...
- Relative exon affinities and suboptimal spli...
- Engagement of L-selectin impairs the actin...
- A radiation hybrid map of the BRCA1 region ...
- [Neuropharmacological studies on drug d...
- Monocyte stimulation by reactive oxygen s...
- Segregation of a missense mutation in the...
- Clinical characteristics of a family with ch...
- Actin dynamics in human neutrophils duri...
- Propofol induces changes in the cytosolic f...

Scientists

1 thru 10 of 158

- Hong M
- Morris JC
- Lee VM
- Trojanowski JQ
- Goate A
- Geschwind DH
- Bird TD
- Reed L
- Schellenberg GD
- Wszolek Z

Organizations

1 thru 10 of 41

- Born Bunge Foundation, Department of Bio...
- Center for Neurodegenerative Disease Re...
- Center for Neurodegenerative Disease Re...
- Center for Neurodegenerative Disease Re...
- Center for Neurodegenerative Disease Re...
- Departamento de Morfologia, Escola Paulis...
- Department of Anaesthesiology and Intensi...
- Department of Anesthesiology, Faculty of H...
- Department of Cell Biology, Linkoping Unive...
- Department of Clinical Genetics, Erasmus ...

Locus

1 thru 1 of 1

PP

Locus

1 thru 1 of 1

Protein Family

1 thru 2 of 2

- Microtubule associated Tau protein family
- Tau protein family

Proteins

1 thru 1 of 1

- * MICROTUBULE-ASSOCIATED PROTEIN TAU

Protein Variants

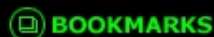
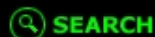
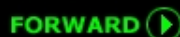
1 thru 10 of 15

- * 275-305 MISSING (variant of MICROTUBULE-...
- * 278-308 MISSING,1-44 MAEPRQEFVMDHA...
- * 45-102 MISSING (variant of MICROTUBULE-...
- * 613-624 MISSING (variant of MICROTUBULE-...
- * ARG406TRP (variant of MICROTUBULE-AS...
- * ASN279LYS (variant of MICROTUBULE-AS...
- * GLY272VAL (variant of MICROTUBULE-AS...
- * GLY389ARG (variant of MICROTUBULE-AS...
- * IVS10, A-G, +13 (variant of MICROTUBULE-...
- * IVS10, C-U, +14 (variant of MICROTUBULE-...

Diseases/Phenotypes

1 thru 4 of 4

- PALLIDOPONTONIGRAL DEGENERATION
- Parkinson's disease
- Pick disease of the brain
- Wihemsen-Lynch Disease



Some types were not visible due to your view settings

Key features of the Knowledge Discovery Tool include dynamic discovery of relationships between datasets and support of collaboration among scientific and institutional experts.



- The user sees data in his informational context. He is able to navigate from the starting point to related items and gain new insights on data.



Further Discovery



- At any point during navigation, the user can refer to a list of Scientists and Organizations that have an expertise in the subject of his interest



Further Collaboration



Questions?