



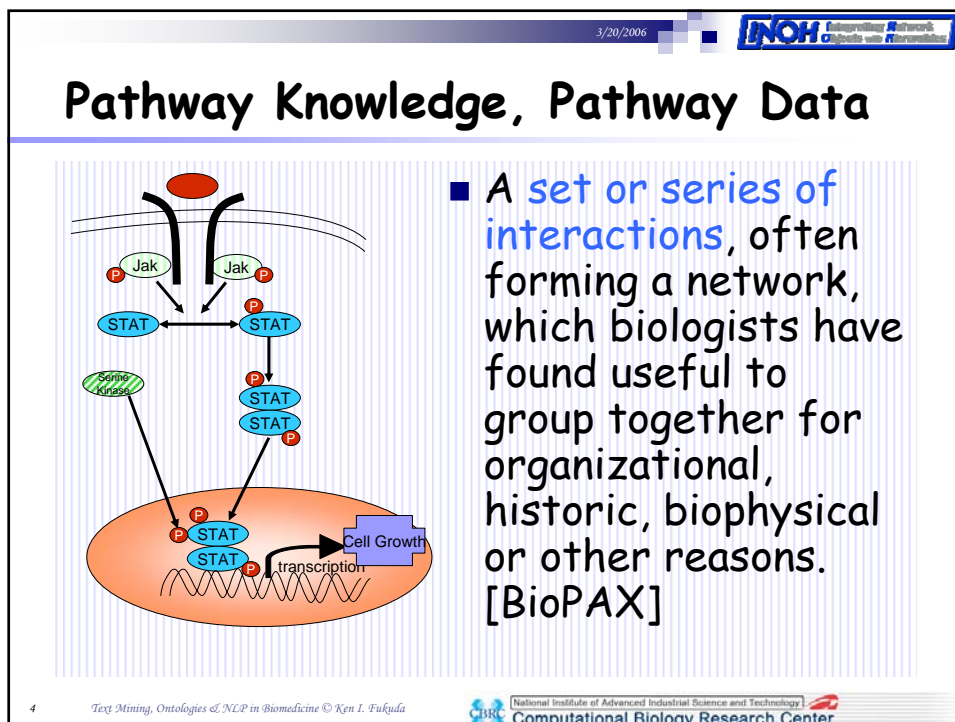
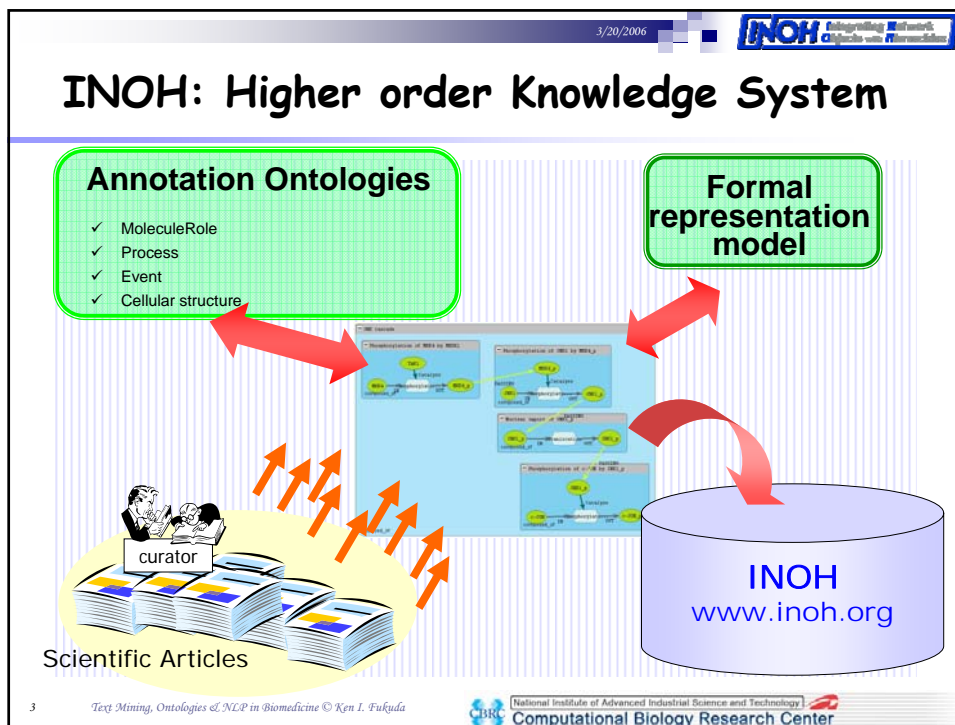
**INOH Pathway Database**  
- Pathway Data and Annotation Ontologies -  
<http://www.inoh.org>

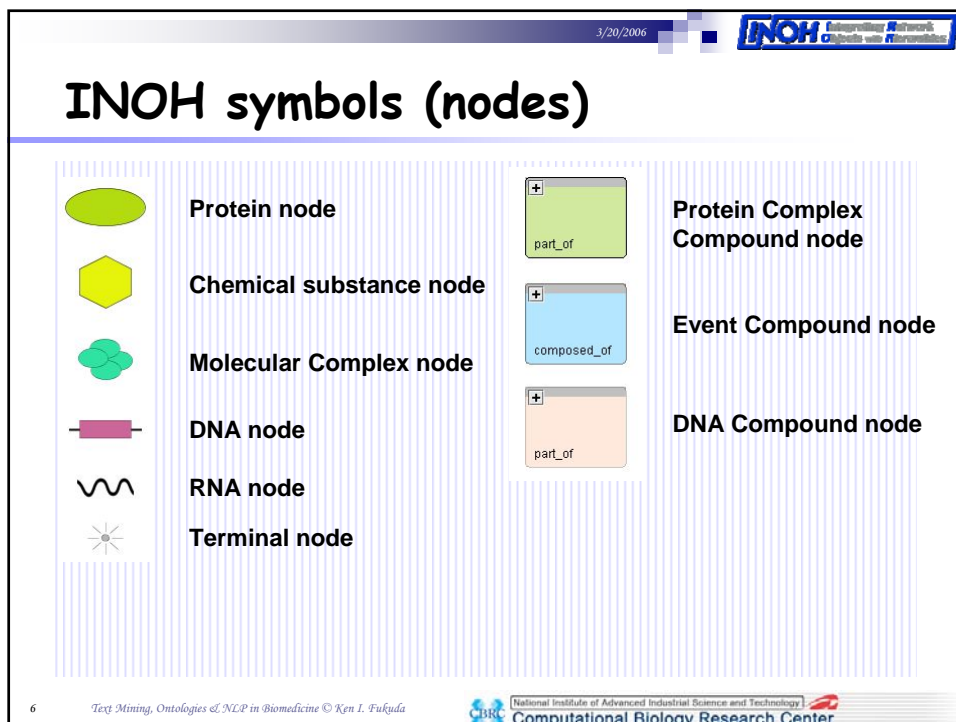
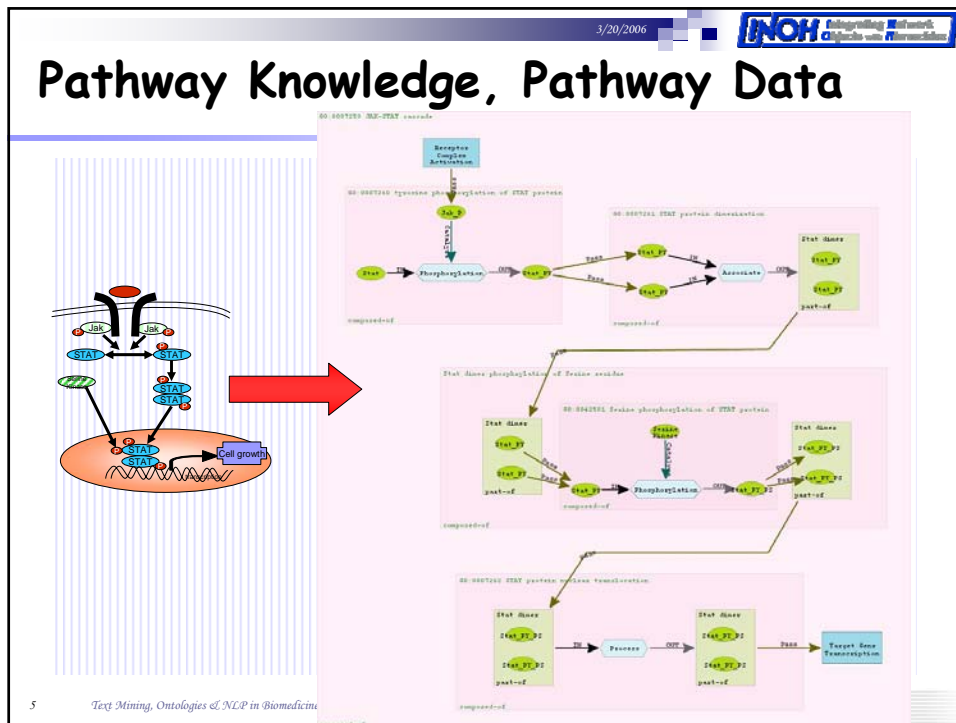



Ken I Fukuda  
CBRC, AIST  
Computational Biology Research Center  
National Institute of  
Advanced Industrial Science and Technology



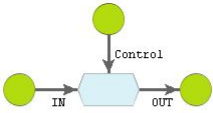
**The INOH DB Project**  
<http://www.inoh.org>



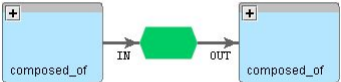


3/20/2006 

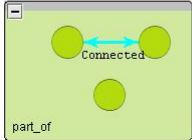
## INOH symbols (edges)





**IN/OUT/Control edge (for interactions)**




**IN/OUT edge (for Event Relation)**

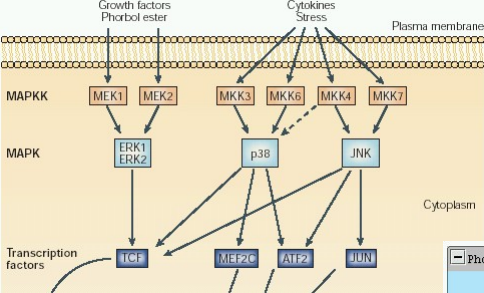


**Connected edge (for additional annotation)**

7 Text Mining, Ontologies & NLP in Biomedicine © Ken I. Fukuda  National Institute of Advanced Industrial Science and Technology 

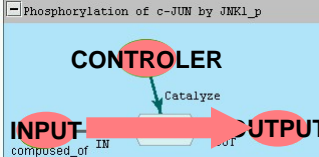
3/20/2006 



## Pathway data ← Graphical Syntax



Growth factors: Phorbol ester  
Cytokines: Stress  
Plasma membrane  
MAPKK: MEK1, MEK2, MKK3, MKK6, MKK4, MKK7  
MAPK: ERK1, ERK2, p38, JNK  
Cytoplasm  
Transcription factors: ATF2, JUN

Phosphorylation of c-JUN by JNK1\_p



8 Text Mining, Ontologies & NLP in Biomedicine © Ken I. Fukuda  National Institute of Advanced Industrial Science and Technology 

### Pathway data <- Graphical Syntax

The diagram illustrates a signaling pathway and its graphical syntax. On the left, a biological pathway shows Growth factors (Phorbol ester) and Cytokines (Stress) interacting with the Plasma membrane. This activates MAPKK (MEK1, MEK2, MKK3) and MAPK (ERK1, ERK2, p3). These lead to Transcription factors (ICF, MEF2c). On the right, a graphical syntax tree for the 'JNK cascade' is shown. It details the phosphorylation of MSK4 by MEK1, the phosphorylation of JNK1 by MSK4\_p, the nuclear import of JNK1\_p, and the phosphorylation of c-JUN by JNK1\_p. Each step is represented as a box with 'Catalyze' and 'Phosphorylates' relationships, and 'composed\_of' relationships between components.

9 Text Mining, Ontologies & NLP in Biomedicine © Ken I. Fukuda  
National Institute of Advanced Industrial Science and Technology  
Computational Biology Research Center

### Pathway data <- Graphical Syntax

This diagram shows the graphical syntax of a pathway as a network of interconnected boxes. Each box represents a step in the pathway, such as 'Phosphorylation of MSK4 by MEK1'. The boxes are connected by lines, representing the flow of information and the relationships between different components. A detailed view of one box at the bottom shows the internal structure with 'Catalyze' and 'Phosphorylates' relationships, and 'composed\_of' relationships between components like c-JUN and c-JUN\_p.

10 Text Mining, Ontologies & NLP in Biomedicine © Ken I. Fukuda  
National Institute of Advanced Industrial Science and Technology  
Computational Biology Research Center

## Standard pathway data syntax

- Biological Pathway Exchange (BioPAX)
  - OWL based data exchange format
  - Community standard for pathway data exchange
  - Entity class
    - <-is\_a- Pathway class
    - <-is\_a- Interaction class
    - <-is\_a- PhysicalEntity class

But data integration is still a hard task.  
Because there are no standards for data value annotation

3/20/2006

Text Mining, Ontologies & NLP in  
Biomedicine © Ken I. Fukuda

11



**Problems with open  
vocabulary annotation**

### Examples from BioCarta

**Concrete Protein**  
"H-Ras"

**Generic Protein**  
"Ras"

Ras  
K-Ras  
H-RAS  
N-RAS  
K-RAS2A  
K-RAS2B

13 Text Mining, Ontologies & NLP in Biomedicine © Ken I. Fukuda  
National Institute of Advanced Industrial Science and Technology  
Computational Biology Research Center

### Examples from STKE


**ERK1/ERK2 MAPK Pathway**

"ERK1" "ERK2"

**G alpha 12 Pathway**

"ERK"

and Technology  
earch Center

3/20/2006 



## Inconsistency in reaction name


- Inconsistency in naming reactions
  - e.g., DBs use the following terms to represent translocation of molecules
    - "translocation,"
    - "transport,"
    - "internalization"

Typically these are open vocabularies and the definition of each term and relations among them are not recorded in DBs

- e.g.
  - "Translocation of BIM to mitochondria [Homo sapiens]"
  - "Notch 2 precursor transport to golgi [Homo sapiens]"
  - "Internalisation of the insulin receptor [Homo sapiens]"

This makes it difficult for the system to retrieve all *translocation related* reactions.

15 Text Mining, Ontologies & NLP in Biomedicine © Ken I. Fukuda  National Institute of Advanced Industrial Science and Technology  Computational Biology Research Center



3/20/2006 

## Inconsistency in reaction name


- Inconsistency in naming reactions

Integration and consistency problem:

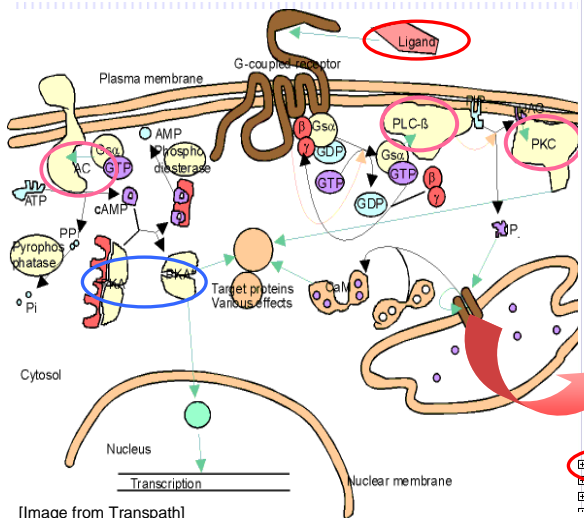
- 1) Computer needs ontologies to bridge the gap between "generic names" and "concrete names"
- 2) difficult to map generic names from different databases
- 3) Difficult to integrate data described in different granularities
- 4) How to assure completeness in querying

16 Text Mining, Ontologies & NLP in Biomedicine © Ken I. Fukuda  National Institute of Advanced Industrial Science and Technology  Computational Biology Research Center

# MoleculeRole Ontology - a protein ontology -

3/20/2006 

## Molecule names are arranged in hierarchical structure to manage the relations between them



- cyclase
- GTPase
- kinase
- protein kinase
  - protein serine/threonine/tyrosine kinase
    - PKA
      - PKA catalytic subunit
      - PKA regulatory subunit
    - PKC
      - receptor serine/threonine kinase
      - protein serine/threonine/tyrosine kinase
      - protein tyrosine kinase
  - PtdIns kinases
  - peptidylprolyl isomerase
  - phosphatase
  - phosphodiesterase
  - phospholipase
  - PLA2
  - PLC
    - PLCbeta
      - PLCbeta1
      - PLCbeta2
      - PLCbeta3
      - PLCbetagamma
  - PLD
  - protease
  - protein related enzyme
  - ligand
  - protein transport protein
  - receptor
  - ...



**Molecular-Complex** (blue circle around PKA)

**Generic Molecule Names** (pink text)

**Concrete Molecule** (pink circle around PLCbeta)

**Molecular Family Names Grouped by Roles** (red circle around ligand)

[Image from Transpath]

18 Text Mining, Ontologies & NLP in Biomedicine © Ken I. Fukuda  National Institute of Advanced Industrial Science and Technology  Computational Biology Research Center


### Concept Relations in MoleculeRole Ontology

The screenshot shows the DAO-Edit interface with a tree view of terms. Red annotations highlight three types of relationships:

- (1) relations between generic molecule names (e.g. ERK) and their concrete molecule names (e.g. ERK1).
- (2) relations between molecular function names (protein serine/threonine kinase) and generic molecule names.
- (3) relations between a complex and its subunits.



Blue and pink callouts on the right side of the screenshot identify "is-a" and "part-of" relationships.

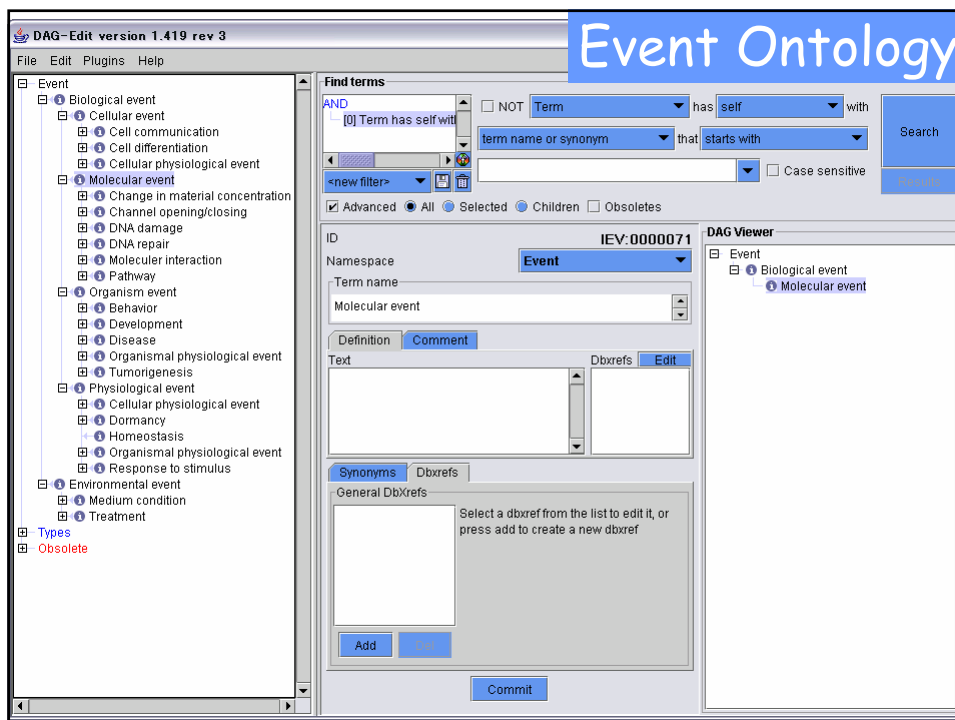
## Event Ontology - a pathway ontology -

3/20/2006 

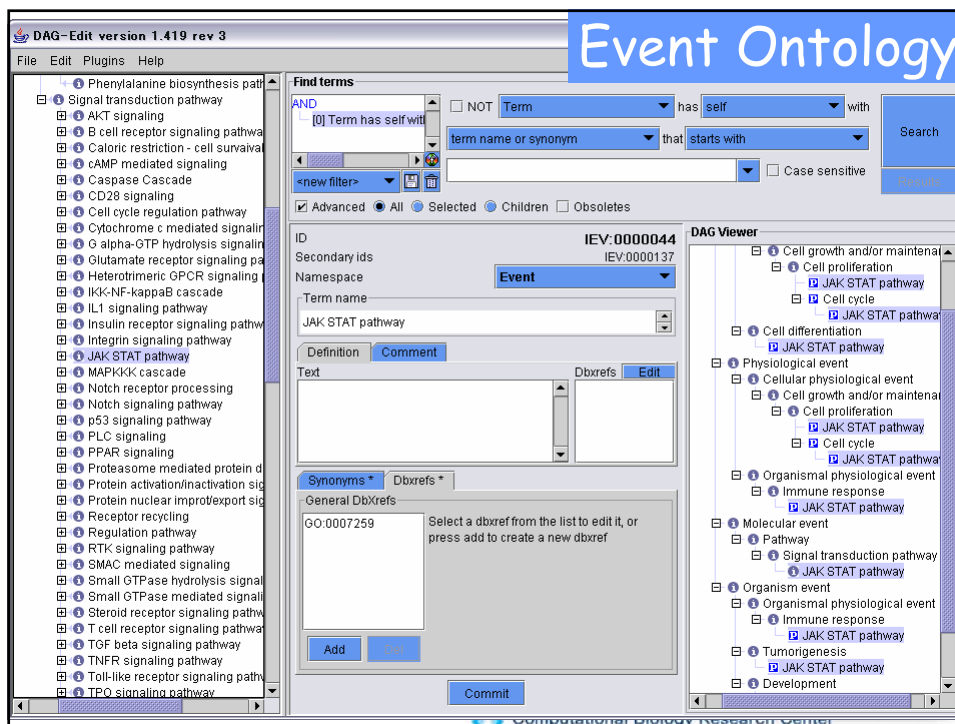
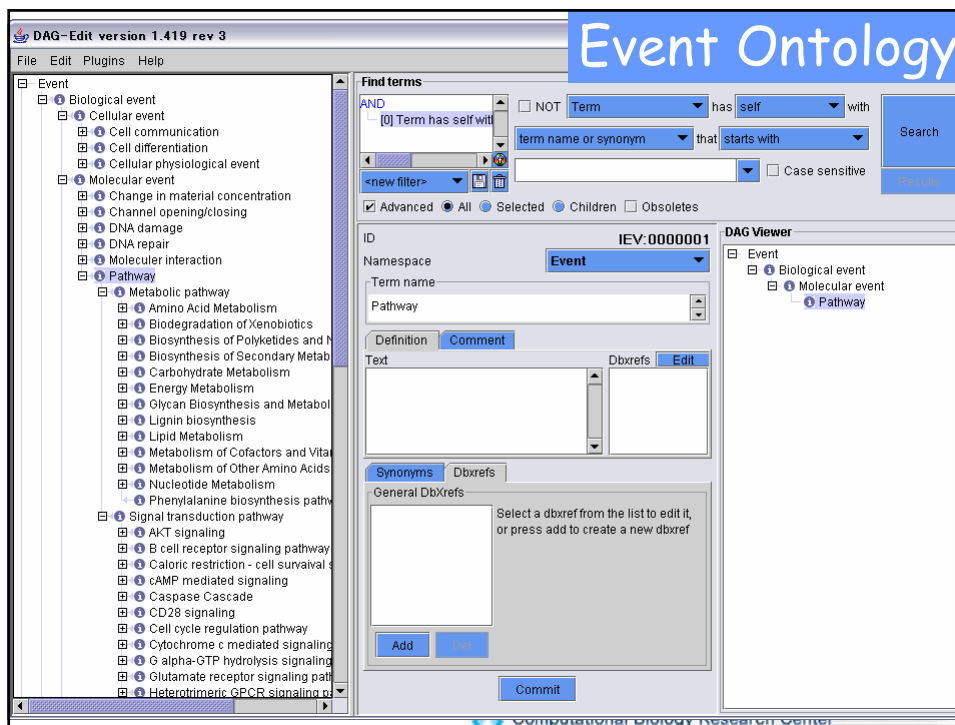
## The INOH Event Ontology

- A DAG structured CV.
- A pathway centric complement to the GO biological-process ontology
- Developed following GO's design convention
  - Naming of concepts
  - how to structure negative/positive regulation pathways
  - etc.
- A paper in the Proc. of PSB2006.

21 Text Mining, Ontologies & NLP in Biomedicine © Ken I. Fukuda  National Institute of Advanced Industrial Science and Technology  Computational Biology Research Center



Computational Biology Research Center




3/20/2006

## Statistics (version 2.10. 2005/Dec/27)



MoleculeRole, total	7,184
protein, total	7,098
chemical, total	85
Cross References	5,076
UniProt ACs	4,582
GO IDs	314
InterPro ACs	31
KEGG COMPOUND IDs	85
PubMed IDs	64


26 Text Mining, Ontologies & NLP in Biomedicine © Ken I. Fukuda National Institute of Advanced Industrial Science and Technology Computational Biology Research Center

3/20/2006 

## Statistics (version 1.59 2006/Feb/08)

Event, total	2,665
KEGG	173
GO	253
Reactome	110

27 Text Mining, Ontologies & NLP in Biomedicine © Ken I. Fukuda  National Institute of Advanced Industrial Science and Technology  Computational Biology Research Center

3/20/2006 

[Main](#) [Criteria](#) [Ontologies](#) [Browse](#) [Project](#) [CVS](#) [Subscribe](#) [Contact](#)

# OBO

open biomedical ontologies

**Open Biomedical Ontologies is an umbrella web address for well-structured controlled vocabularies for shared use across different biological and medical domains.**

This site contains ontologies and points to some other efforts within the community. Ideally we see a range of ontologies being designed for biomedical domains. Some of these will be generic and apply across all organisms and others will be more restricted in scope, for example to specific taxonomic groups.

- View the OBO ontologies in table form
- Browse the OBO ontologies
- OBO news
- Links to related projects



---


**News**

**Gene Ontology survey**

The Gene Ontology survey is now open. It's a short set of questions, which should take a maximum of 10 minutes to complete, and we would be grateful for all responses.



**Human pathways database survey**


28 Text Mining, Ontologies & NLP in Biomedicine © Ken I. Fukuda  National Institute of Advanced Industrial Science and Technology  Computational Biology Research Center

3/20/2006 

## INOH Annotation Ontologies

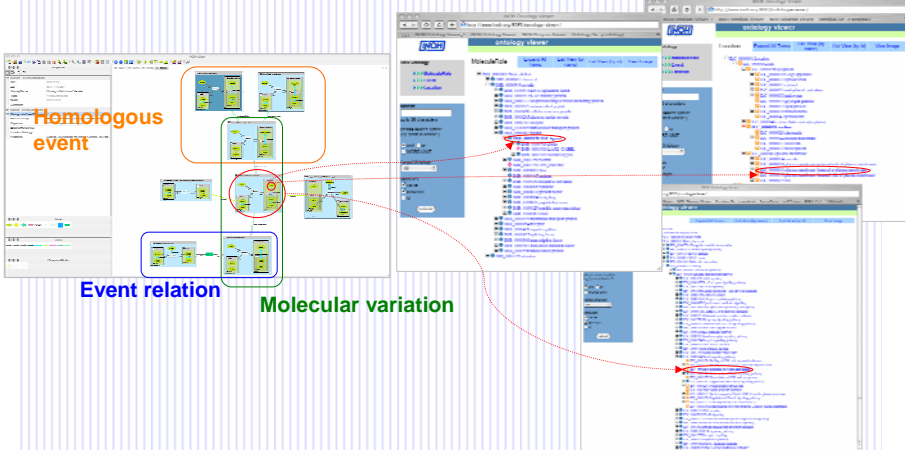
- Available from
  - <http://www.inoh.org/>
  - <http://www.inoh.org/ontology-viewer/>
- Protein Ontology
  - "MoleculeRole Ontology", 7,184 terms
  - Also available from the OBO site
- Pathway Ontology
  - "Event Ontology", 2,665 terms
  - Also available from the OBO site
- Interaction Ontology
  - "Process Ontology" 42 terms
  - 100% mapping with PSI-MI
- Cellular Localization
  - "Location Ontology", 49 terms
- Pathway Relation
  - "EventRelation Ontology", 12 terms
- DNA Sequence Ontology
  - "GenomeSequence Ontology", 57 terms

29 Text Mining, Ontologies & NLP in Biomedicine © Ken I. Fukuda  National Institute of Advanced Industrial Science and Technology  Computational Biology Research Center

3/20/2006 

## INOH pathway data annotation



• Every pathway object is annotated by a set of ontologies

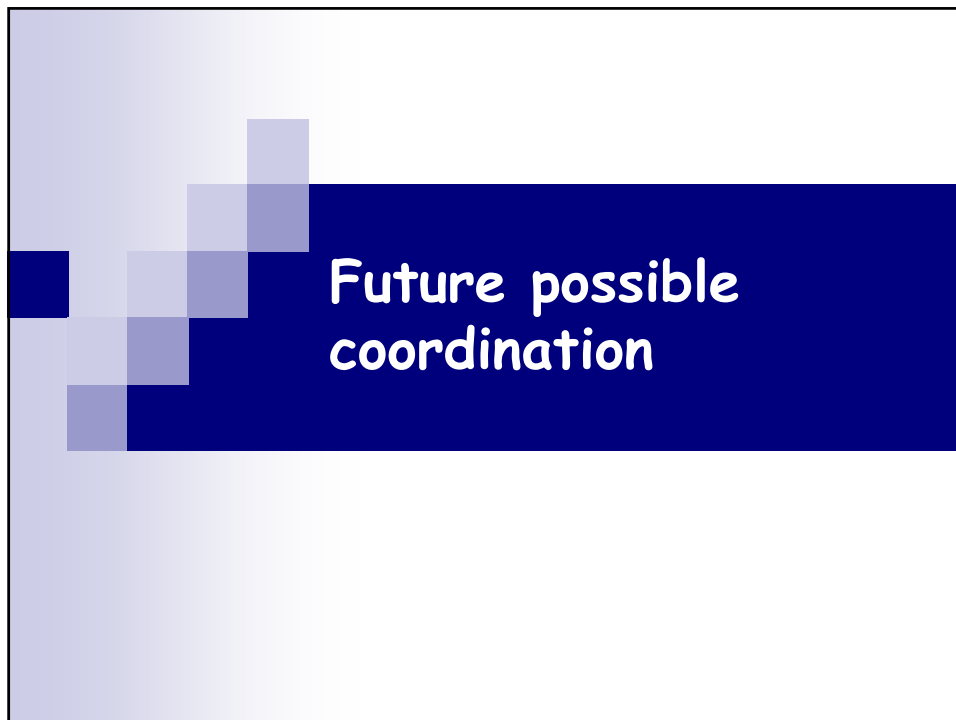



Homologous event

Event relation

Molecular variation

30 Text Mining, Ontologies & NLP in Biomedicine © Ken I. Fukuda  National Institute of Advanced Industrial Science and Technology  Computational Biology Research Center



3/20/2006 

## Event ontology: Compositional structure

### Parsing, decomposing terms

**kind-of**

Binding in cytosol (IEV:0000013)



Binding (IEV:0000004) **Process Ontology term**


**located-in**

Phosphorylation in cytosol (IEV:0000025)

cytosol(ILC:000015) **Location Ontology term**

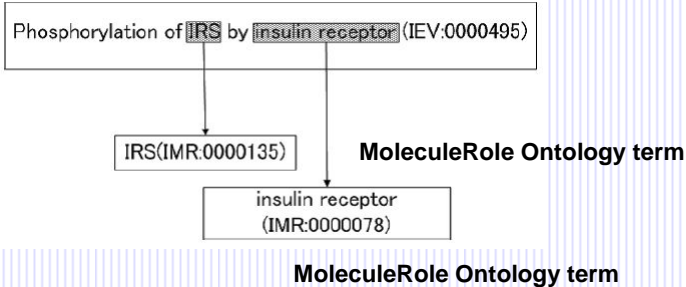
By SUGAI, Tomoyuki (Waseda Univ.)

32 *Text Mining, Ontologies & NLP in Biomedicine* © Ken I. Fukuda  National Institute of Advanced Industrial Science and Technology  Computational Biology Research Center



3/20/2006 

## Event ontology: Compositional structure Parsing, decomposing terms

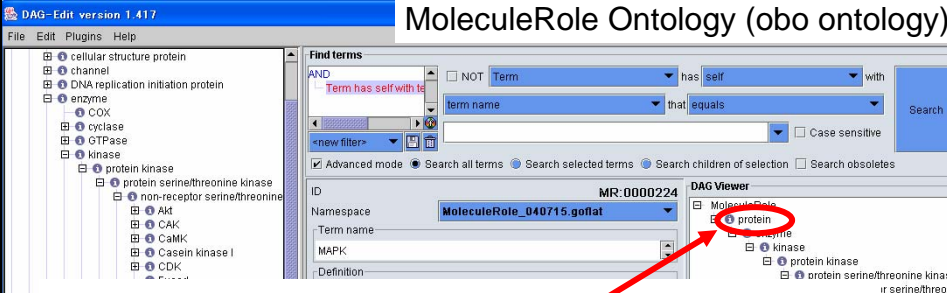
**has-participant**  
**has-agent**

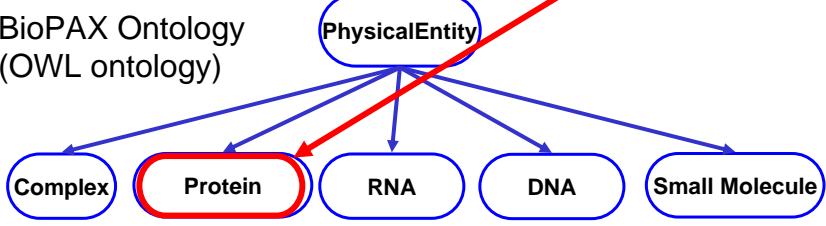


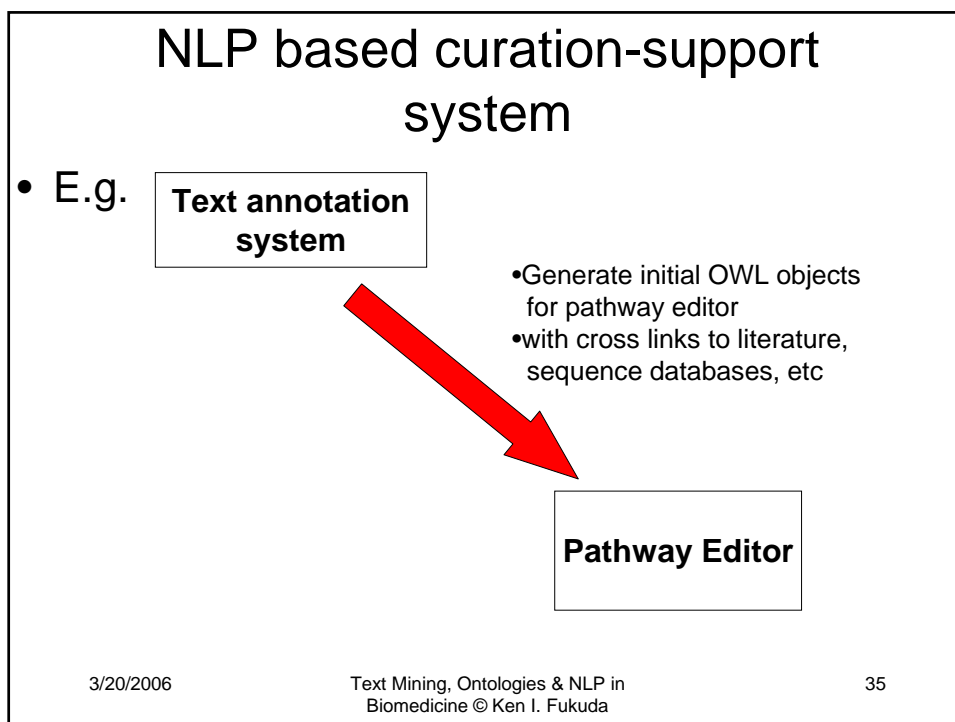
By SUGAI, Tomoyuki, (Waseda Univ.)

33 Text Mining, Ontologies & NLP in Biomedicine © Ken I. Fukuda  National Institute of Advanced Industrial Science and Technology  Computational Biology Research Center

## Combining obo ontology and OWL ontology







3/20/2006

## INOH database team

Toshihisa Takagi, Univ. of Tokyo (PI)  
Ken I Fukuda, CBRC, AIST (Co-PI)  
Takao Asanuma, JST (System Engineer)  
Emi Hattori, IMS (System Engineer)  
Hiromi Nakamura, IMS (System Engineer)  
Tatsuya Kushida, JST (Event Ontology Editor, curator)  
Satoko Yamamoto, JST (MolRol Ontology Editor, curator)  
Yuki Yamagata, JST (curator)  
Noriko Sakai, JST (curator)  
Yuki Nakanishi, Intec W&G (curator)

### Funding

BIRD, JST  
A Grant-in-aid for Scientific Research on Priority Areas, "Genome Science", from MEXT

36 Text Mining, Ontologies & NLP in Biomedicine © Ken I. Fukuda  National Institute of Advanced Industrial Science and Technology  
Computational Biology Research Center