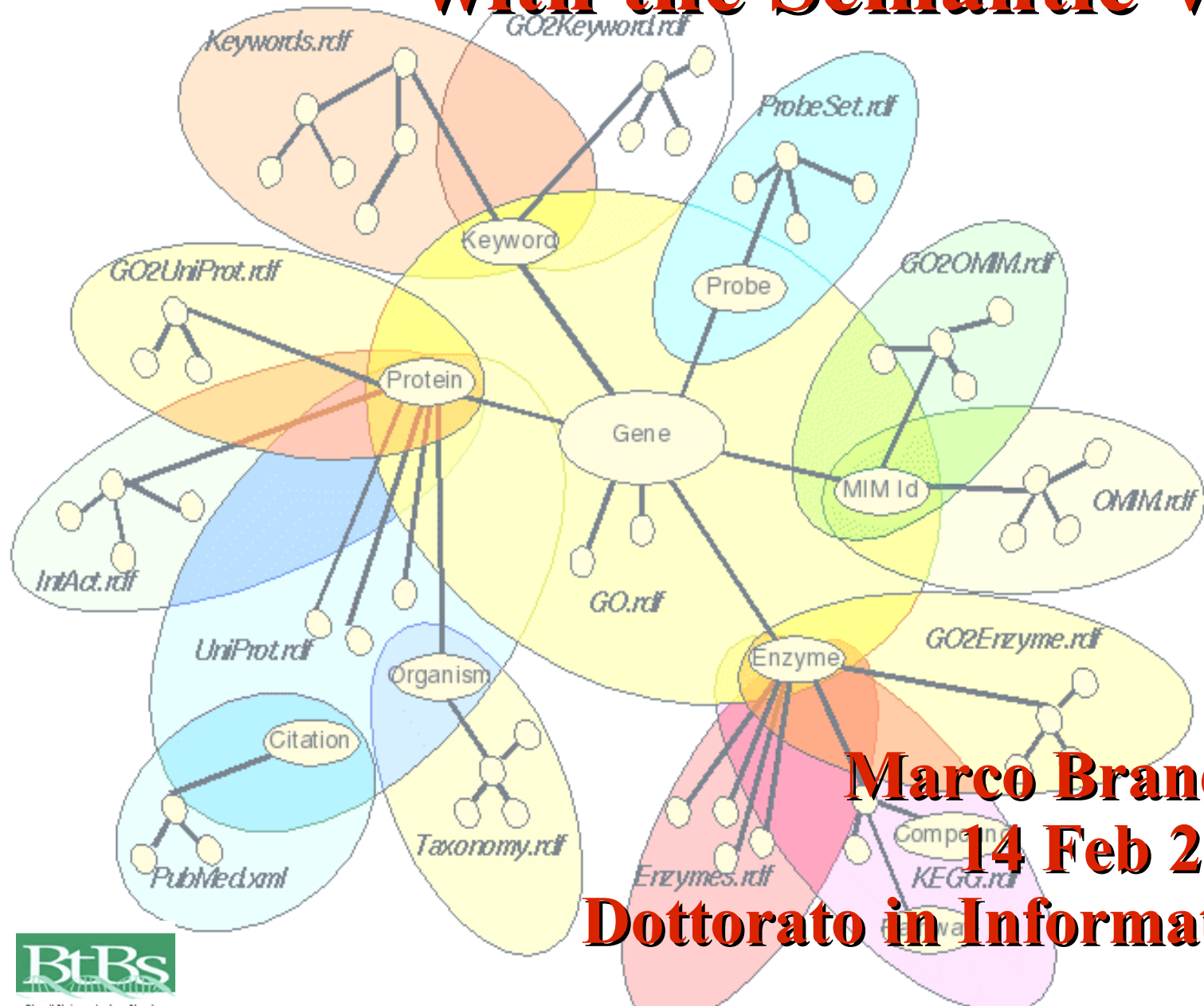


Managing Microarray Knowledge with the Semantic Web



Marco Brandizi
14 Feb 2008
Dottorato in Informatica

Microarrays and Gene Expression Analysis

Formal models in Microarray Knowledge

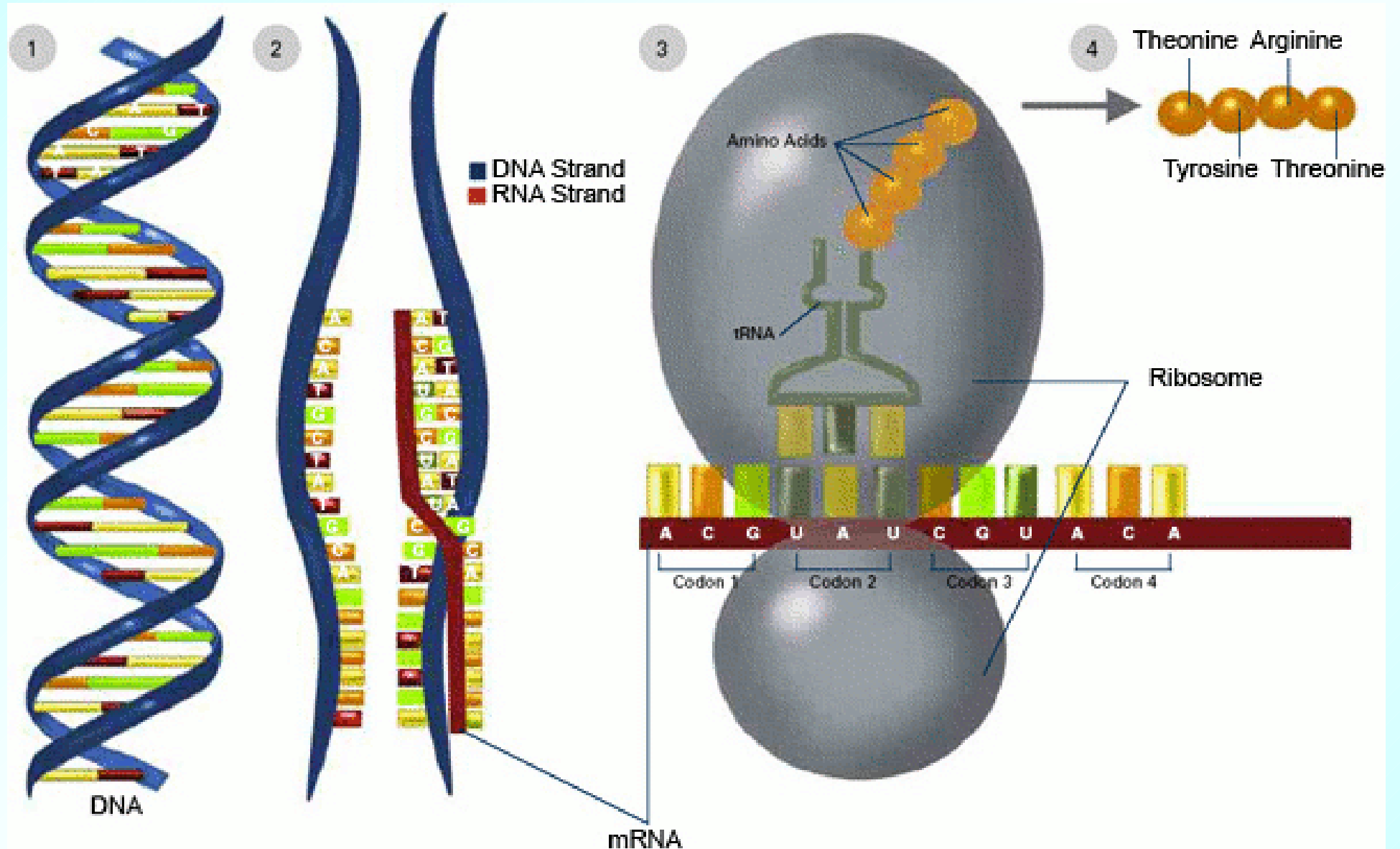
Our Semantic Web based proposal

A demo application

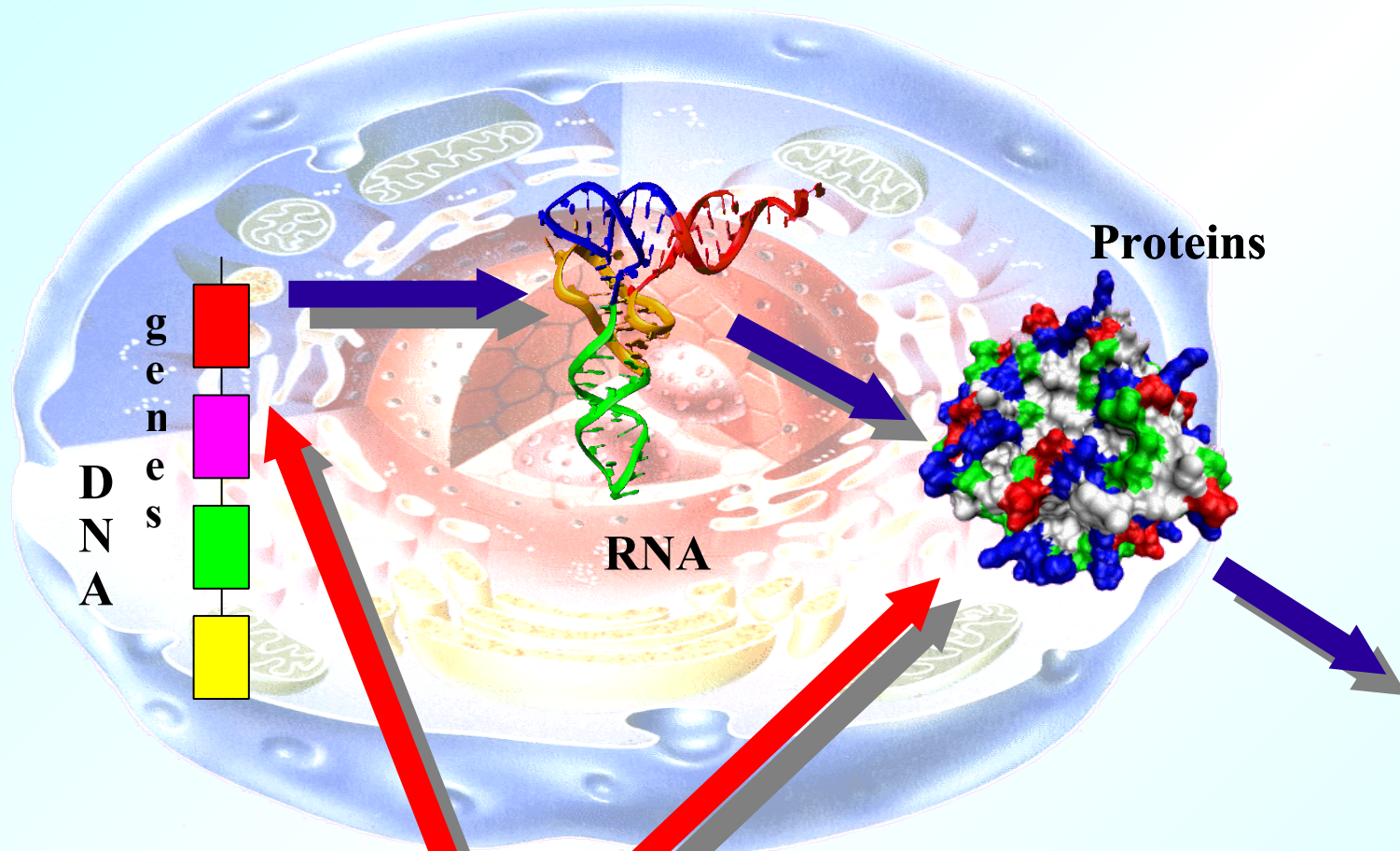
A proposal for Knowledge Ranking

Conclusions/Future

Gene Expression



Gene Expression



by Alessandro C. (Studio Stalio) - copyright 2000 Andrea due

Gene Expression

Life

Why studying Gene Expression?

- **Functional Genomics:**

- What is the **function of these genes** (similar expression profiles)?
- Which genes are involved in Immune response?

- **Genetics studies**

- Which **gene mutations** characterise this disease?

- **System Biology**

- Which **pathways** are activated in during cell cycle stages?
- Which network **modules** explain gene-gene correlation?

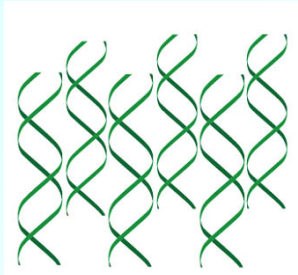
- **Diagnostics**

- A given expression profile may characterise cancer cells, early diagnosis possible

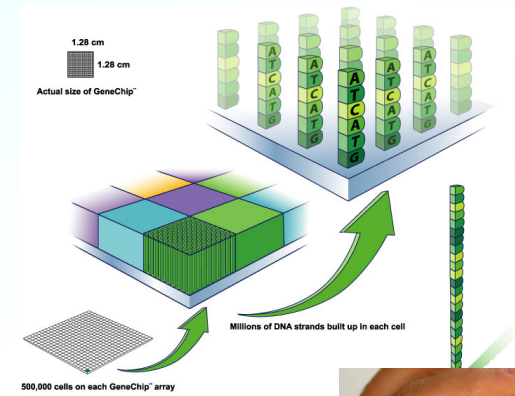
Microarray Technology



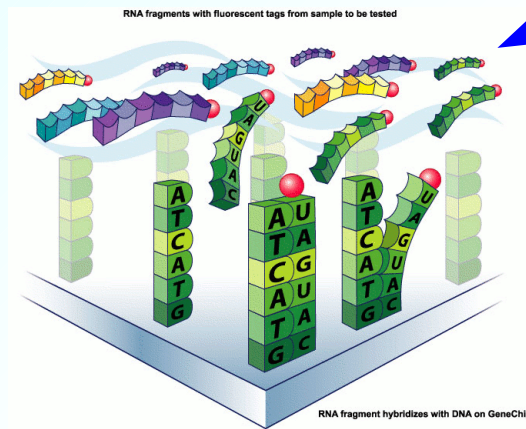
Organism
(control+treated)



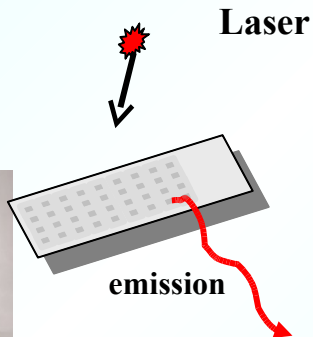
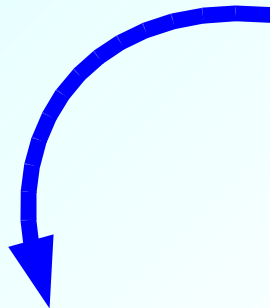
RNA fragments, labelling



Microarray Chip

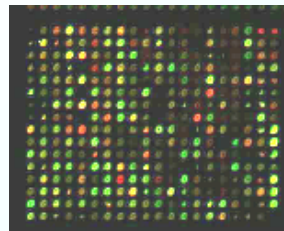


Hybridization

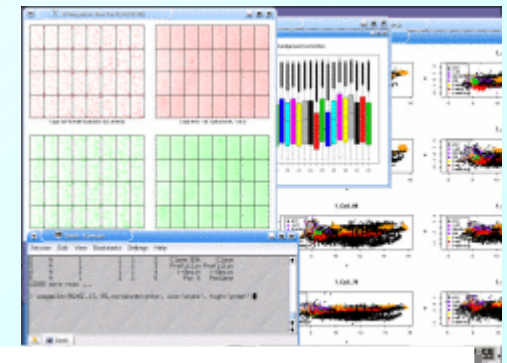


Laser

emission



Scanning/Measurement



Data, Analysis,
New biological knowledge

Why Microarrays?

- It's an **High Throughput Technology**
- **Thousands of measures**, about thousands of gene, in one single measurement operation (hybridization)
- **Olistic** approach: we may find insights analysing many factors (genes) at the same time, not gene-by-gene
- Systems biology: we may model the whole system
- **Bioinformatics and Statistics methods**
 - Statistical analysis
 - simulation and modelling (e.g. Systems Theory)
 - Data may be publicly distributable and analysed many times, by many people
- May be **integrated with other HT technologies** (e.g.: Mass Spectrometry)
- May be integrated with **wet lab approaches**

How to represent Microarray Data: MIAME/MAGE Standards

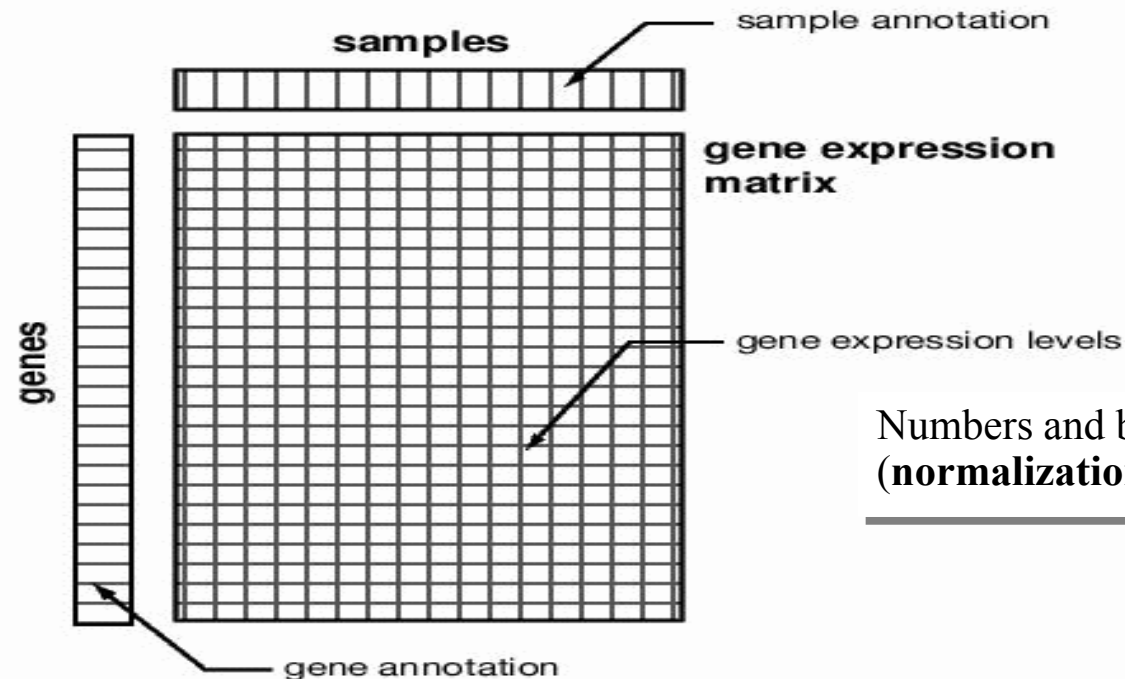
Samples are grouped into **experiments**

- Experiment meta-data (title, date, objective)
- Experimental **Factors** (independent variables)
- **Authors** and Derived **Publications**

- **Organism** and **characteristics** (age, sex...)
- **Treatments** (compound, infection, dose...)
- Preparation and methods (**protocols**, **labelling** substance...)

(Mostly Pub Sources)

- **Gene** Names
- Gene Products (**Proteins**)
- Gene/Proteins Biological **Pathways** and interactions
- Gene/Protein **Sequence**
- Other kind of RNA/DNA sequences



Numbers and basic processing
(**normalization**, **scaling**)

Array Design (which sequences are in the device, where, grouping of probes)

Why Standard formats?

- **Basically**: experimental activity has to be **understandable** by the whole scientific community and **reproducible**
- **Comparable** results is also very valuable
- For instance:
 - we may want to compare the expression of a set of known genes, under different conditions, using data from different laboratories
 - **genes** = standards for gene representation (e.g.: Unigene) and annotations (e.g. GeneOntology)
 - compare the **expression** = description of data processing and data production process
 - different **conditions** = terminology for factor types, experimental design
 - different labs = standard **meta-data for experiments**.
- We want redo an experiment to check its reproducibility
 - precise description of **experimental design, biological materials, laboratory protocols** and the **pipeline** that has lead from the organisms to the final data.
 - description/reference of the **arrays** that have been employed

Microarrays and Gene Expression Analysis

Formal models in Microarray Knowledge

Our Semantic Web based proposal

A demo application

A proposal for Knowledge Ranking

Conclusions/Future

How is the Transcriptomics Analysis Output?

- **Set of Genes**
 - List of **Differentially Expressed Genes** (Control vs. Conditions)
- **Set of Clusters** (Hierarchical Gene Sets)
 - Bi-Clusters (Both genes and conditions are grouped into clusters)
- **Enriched Set of Genes**
 - Sets are classified according to Experimental Factors and/or to biological categories (e.g.: to Gene Ontology or to KEGG pathways)
 - Experimental factor = independent variable that is varied to study the effect of its variation on the gene expression
- More abstract Knowledge
 - **Biological Conclusions or insights**
 - **Papers, Books**
 - Knowledge about materials and methods (often tacit)

And behind all above: Biological Knowledge

Examples of Microarray Knowledge

● Genes, Gene products

- Under which conditions (exp. factors) is this gene expressed?
- Which data/experiments do support a transcription profile?
- Are the genes related to a function expressed under X?

● Conditions, Experiments

- Which genes are expressed on diabetic patients?
- Which experiments are about dendritic cells and which genes are mostly expressed?

● Data, Protocols, Methods

- For which organism does this protocol perform best?
- Is this data set, leading to a given conclusion, reliable? Did the experts noted any problem?
- Is a given platform/array more used with certain experimental designs?

● People, knowledge authoritativeness

- Who is studying these genes?
- How many times a claim has been concluded by an analysis? By who? What is his/her role? Does he/she have important publications on the topic?

How to represent Results from Microarray Analysis

Natural Language



More formally



Related Work

• Repositories and LIMS

- Parkinson et al., *ArrayExpress—a public database of microarray experiments and gene expression profiles*. (2007). Ranks genes according to experiments and conditions of expression. Limited overall analysis, no collaboration features.
- Kapushesky et al., *Expression Profiler: next generation-an online platform for analysis of microarray data*. Allows for management of DEG/clustering lists. Limited collaboration.
- (More modestly...) Brandizi, Splendiani et. al., *The Genopolis Database*. Allows data sharing among collaborators groups. Gene lists. Focused on Affymetrix/DCs. Limited inference.

• Collaboration Systems

- **Array Management Manager** by Biodiscovery. Basic analysis features and results sharing. Non standard (web) interface.
- **Synapsia** by Agilent. Narrative, hypothesis driven discovery and collaboration system. Non standard (web) interface.

• Knowledge-based systems

- Massar et al., *BioLingua: a programmable knowledge environment for biologists* (2004). Frame-based system for data integration, non microarray-specific, limited user feedback & collaboration.
- Racunas et al., *HyBrow: a prototype system for computer-aided hypothesis evaluation*. Frame-based system for pathway-based investigation. Limited user feedback and collaboration.

Microarrays and Gene Expression Analysis

Formal models in Microarray Knowledge

Our Semantic Web based proposal

A demo application

A proposal for Knowledge Ranking

Conclusions/Future

Our approach: Microarray-specific modelling with the Semantic Web

- **MANN: an OWL model** for representing experiments, results, people, hypothesis/conclusions (more later)
- **A Web Demo**, based on the Makna Semantic Wiki
 - that shows some real cases
- Examples of pre-crafted **SPARQL queries**, that have biological significance
- A proposal for **ranking OWL-modelled knowledge**

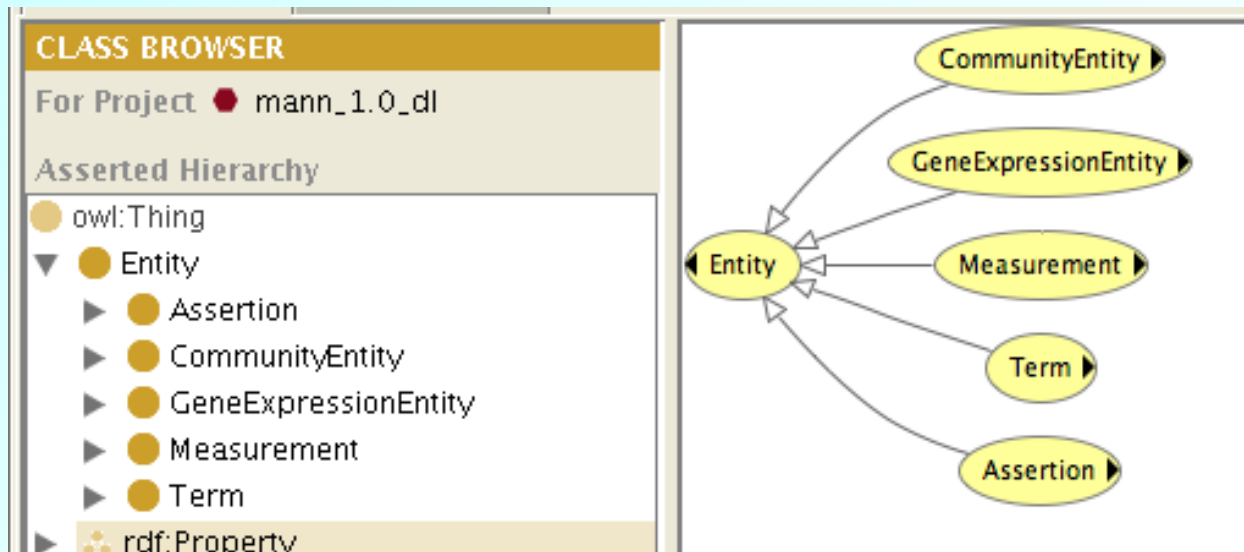
Why The Semantic Web for Biology?

- **Public knowledge** (“resources”) that is **highly interrelated**
- **Wide use of WWW** and growing interest for **RDF integration**
 - LSID project
 - HCLS group
- Very **heterogeneous** models, data types, etc.
- **Ontologies** (or similar models) are much needed in Biology and already extensively used
 - Need to conceptualise
 - Need to standardise the conceptualisations and available information
 - Even basic inference may be useful (and still not much used)
 - Increasing use of OWL and DL
- **Similar projects** already existing that make use of OWL and/or SW in general
 - HCLS Demo
 - ART Ontology, Soldatova et al., 2007
 - Ontology for Biological Investigations (OBI), <http://obi.sourceforge.net/>

The MicroAnnOnto (MANN) OWL Model

- Informally: an Ontology
- Tries to reuse existing models/ontologies (MGED-Ontology, GeneOntology, SKOS, SWRC/COIN)
- Models **Assertions** (e.g.: hypotheses and conclusions) which are supported by Microarray Data, have authors, and more
- Models **Microarray Experiments**, with the aim of linking assertions and data
- Models **Data Evaluation and Ranking**, allowing for user feedback and data quality management (e.g.: precision, significance, likelihood)
- Promotes collaboration: comments, people and roles, papers, conferences

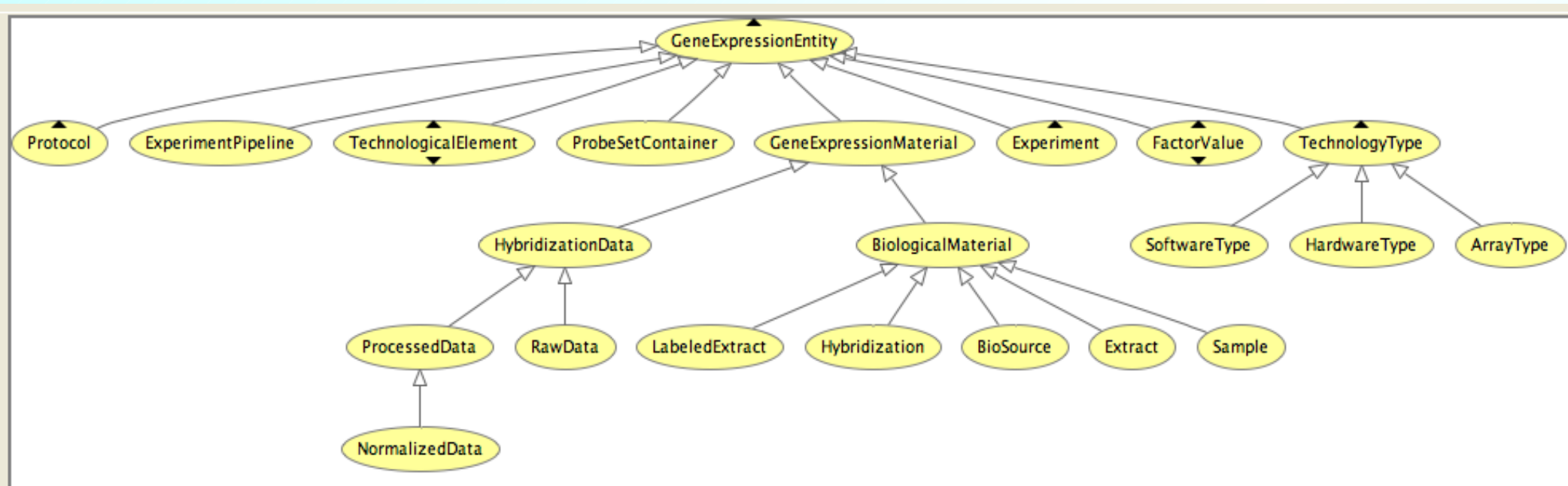
The MANN Model: Basic Classes



● Basic Properties

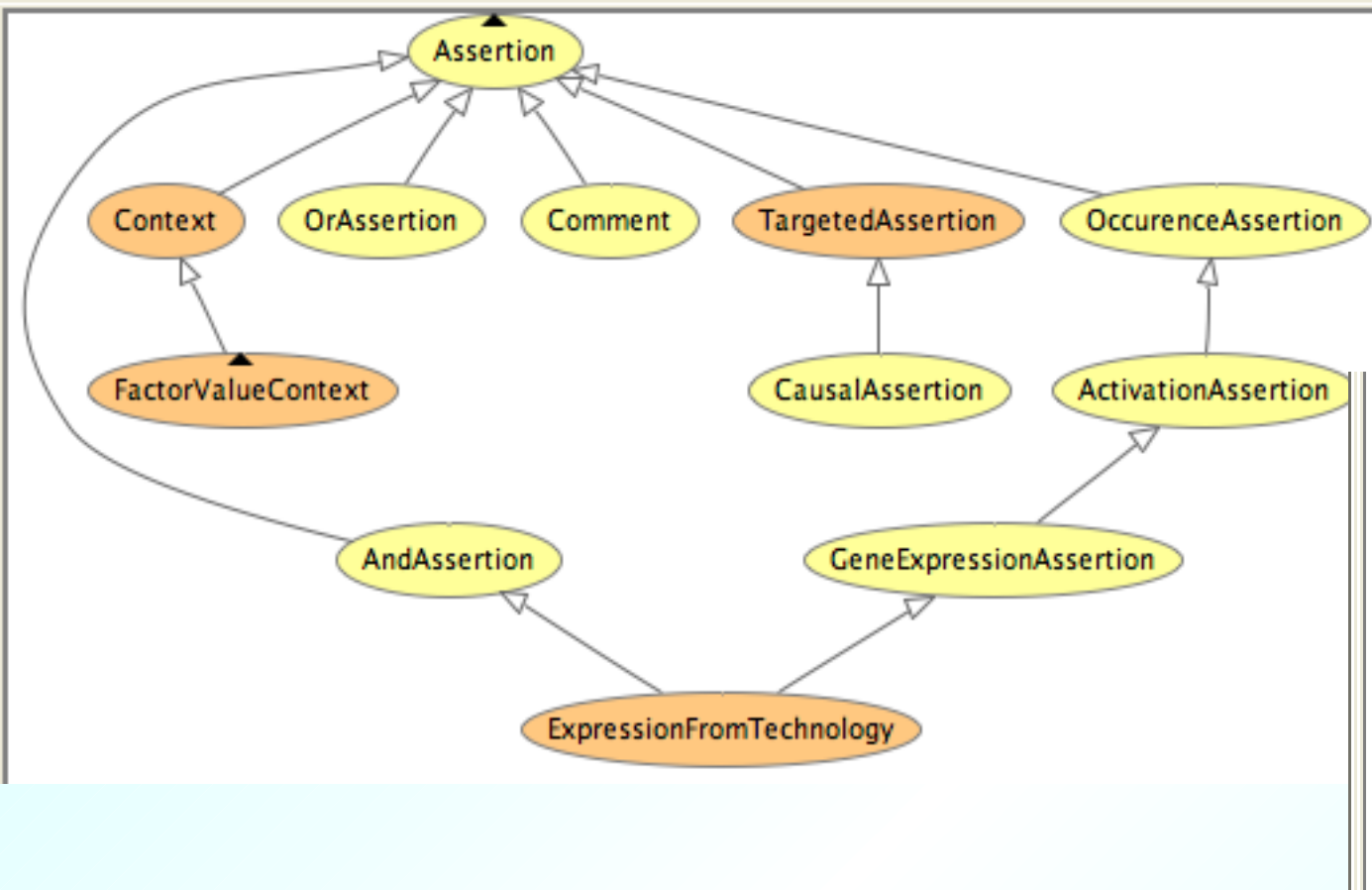
- entityName (short name), entityTitle, entityDescription
- entityOwner(Entity->PeopleEntity)
- **evaluation** (Entity->Literal, typically numbers)
- termAnnotation(Entity->Term)
- entitySupports(Entity->Entity) more in the follow

The MANN Model: Gene Expression Classes²⁰



- GE Properties
- usesGentity(GeneExpressionEntity->GeneExpressionEntity)
- Specific usage (arrayProbes, arrayHasType)

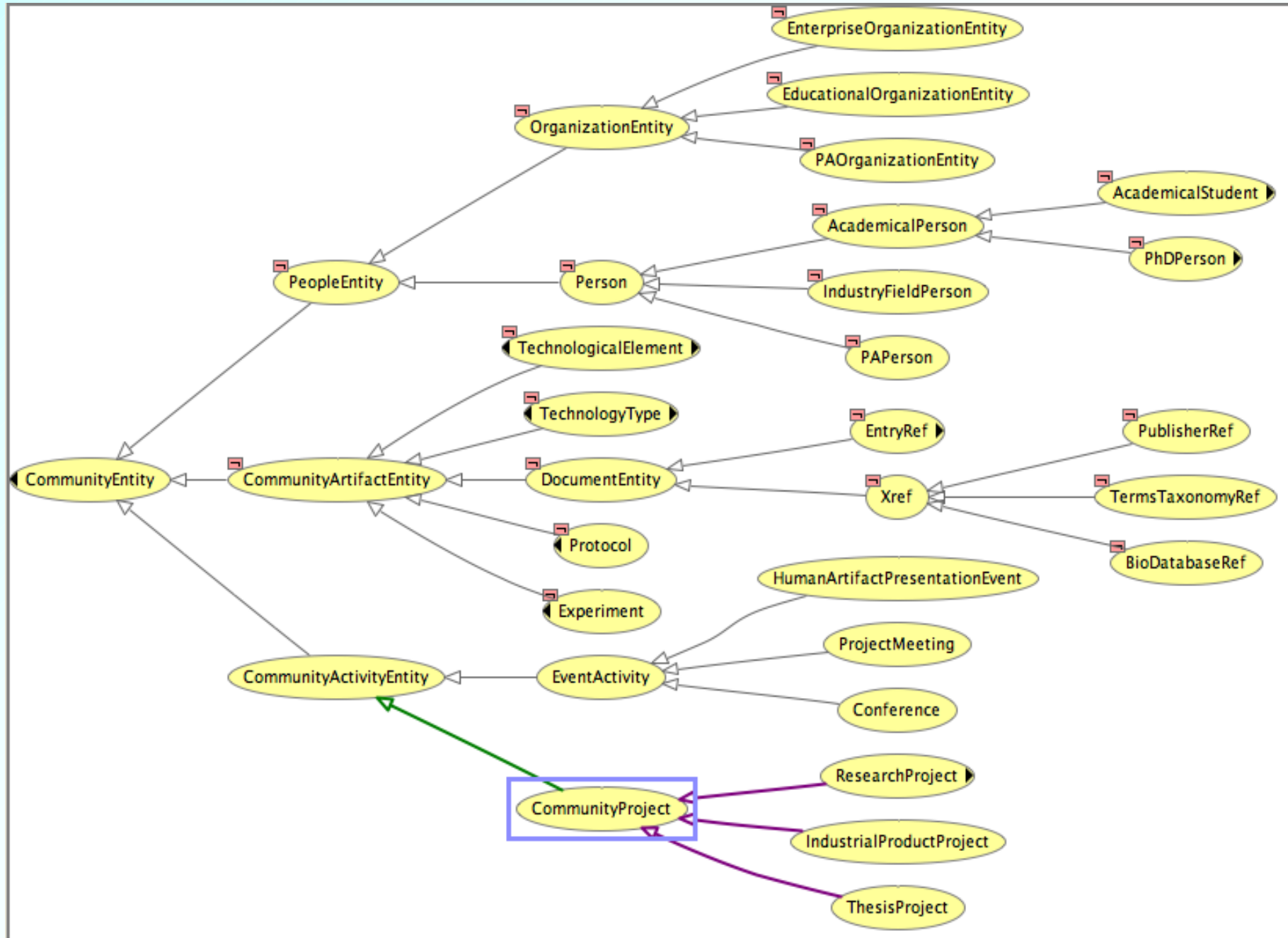
The MANN Model: Assertions



- relatedAssertions ↔ relatedAssertion
- ▼ ■ hasBroaderAssertion ↔ hasNarrowerAssertion
- equivalentAssertion
- ▼ ■ hasNarrowerAssertion ↔ hasBroaderAssertion
- equivalentAssertion
- isComponentOfAssertion ↔ hasComponentAssertion
- hasComponentAssertion ↔ isComponentOfAssertion
- oppositeAssertion
- ▼ ■ similarAssertion
- equivalentAssertion

- Assertions have subjects, which are intended in And combination (e.g.: all true at the same time) or in Or combination
- An assertion may have a context to be referred to (e.g.: Gene X is expressed in leukaemia cells)
- A set of assertion-assertion properties is available
- Supporting data, authors, terms and other attributes may be added

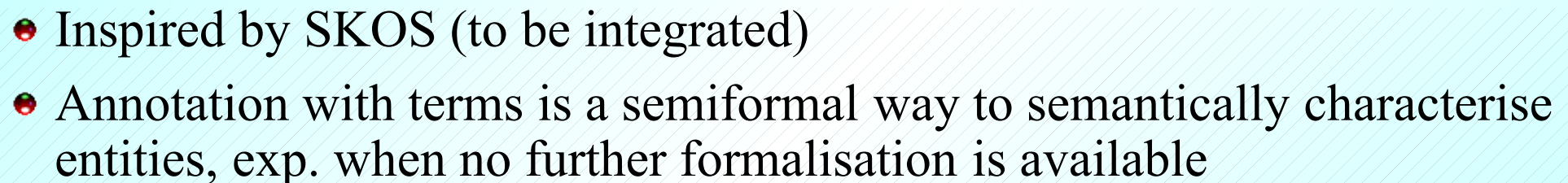
The MANN Model: Community Entities



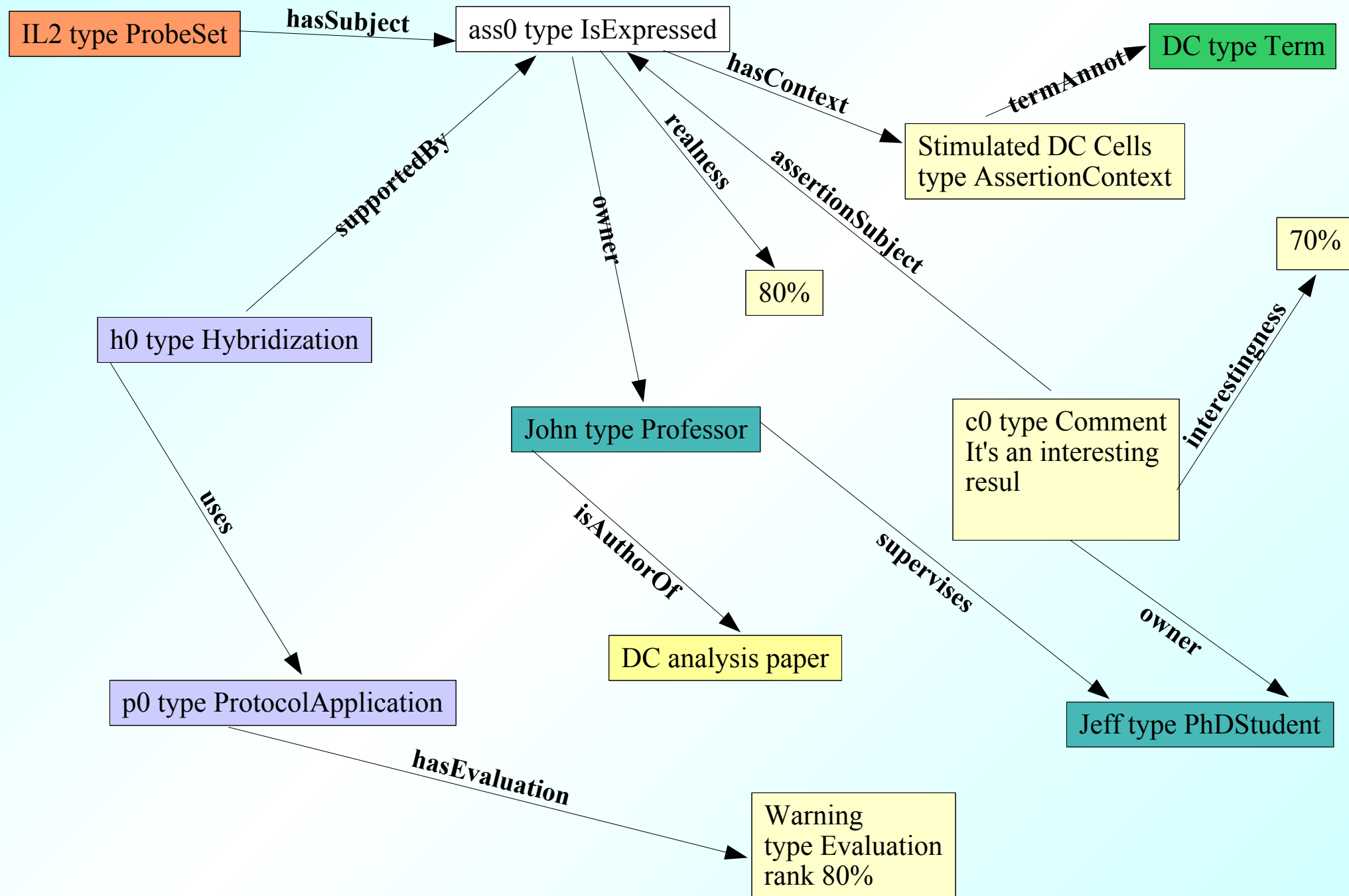
The MANN Model: Community Entities

- Inspired by SWRC+COIN
- To be integrated
- Roles for CommunityEntity(s) may be important in searching and ranking
- e.g.: an assertion stated by a professor could be ranked more than one stated by a PhD student
- For the same reason a set of role-role properties is defined
- Applicability of this part is currently limited by data availability
- Co-citation is one of the available things

- cooperatesWith ↔ cooperatesWith
- organizationHasManager ↔ isManagerOfOrg
- studiesAtOrganization ↔ organizationHasStu
- ▼ ■ isLeaderOf ↔ hasLeader
 - isHeadOf ↔ hasHead
 - isSupervisorOf ↔ isSupervisedBy
 - isCoordinatorOf ↔ coordinatedBy
 - isManagerOfPerson ↔ personHasManag
 - isManagerOfOrganization ↔ organization
- hasHead ↔ isHeadOf
- organizationHasStudent ↔ studiesAtOrganiz
- ▼ ■ isEmployeeOfOrganization ↔ organizationH
 - isManagerOfOrganization ↔ organization
- organizationHasEmployee ↔ isEmployeeOfC
- personHasManager ↔ isManagerOfPerson
- teachesAtOrganization ↔ organizationHasTe
- ■ hasLeader ↔ isLeaderOf
- organizationHasTeacher ↔ teachesAtOrgani
- coordinatedBy ↔ isCoordinatorOf
- ■ hasColleague



The MANN Model: Application example



Microarrays and Gene Expression Analysis

Formal models in Microarray Knowledge

Our Semantic Web based proposal

A demo application

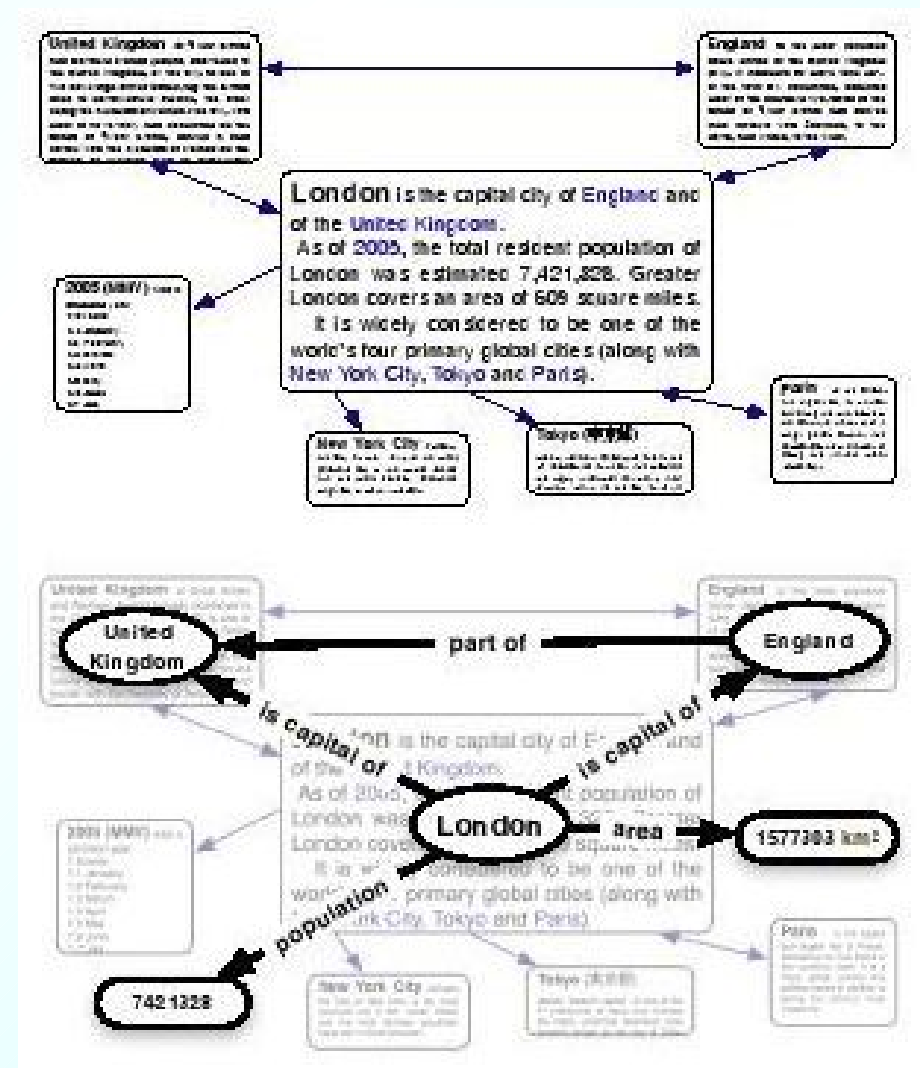
A proposal for Knowledge Ranking

Conclusions/Future

A Demo Application: MannMakna Wiki

The screenshot shows the IkeWiki web application. At the top, there are tabs for 'article', 'context', 'mozile editor', 'edit content', 'edit metadata', 'edit types', 'edit links', and 'discussion'. The main content area is titled 'FrontPage' and includes a 'Contents [show]' button. Below this is a 'Welcome to IkeWiki' section with a paragraph describing the application as a Semantic Wiki developed by SalzburgResearch. To the right, there is a 'references' section with a list of outgoing and incoming references, including 'SalzburgResearch', 'cc:license(1)', 'dc:rights(1)', 'dcterms:license(1)', 'help:hasTableOfContents(1)', 'IkeWiki:hasAuthor(2)', 'user:SebastianSchaffert', 'user:mzsilva', 'rdf:type(3)', 'cc:Work', 'rdfs:Resource', 'wordnet:Document', and 'help:isTableOfContentsFor(1)'.

[<http://ikewiki.salzburgresearch.at/>]



[<http://www.aifb.uni-karlsruhe.de/WBS/hha/papers/SemanticWikipedia.pdf>]

Makna and Jena

- Makna: a Semantic Wiki, based on JSPWiki
 - Has a **simple syntax** for editing contents
 - Has a **simple interface**, where the semantic structure of a page/resource is clearly shown
 - It's based on Jena
- Jena: a comprehensive framework for the Semantic Web, including features like:
 - Persistence with DBMS back-ends (we use MySQL)
 - OWL support, OWL reasoning, either via DIG or with built-in reasoners
 - **Rule system** that allows for adding custom inference
 - SPARQL
- Pro and Cons
 - Makna has limited support for pages with **many statements** (ranking and filtering being developed)
 - Makna badly supports **some SW features** (e.g.: URIs that are not wiki URLs, namespaces, labels, typed literals)
 - **Jena performance** (esp. in reasoning) is not great (optimizations being developed)

An real case stored in MannMakna

A Type I IFN-Dependent Pathway Induced by *Schistosoma mansoni* Eggs in Mouse Myeloid Dendritic Cells Generates an Inflammatory Signature¹

François Trottein,^{2*} Norman Pavelka,^{2†} Caterina Vizzardelli,[†] Veronique Angeli,^{*} Claudia S. Zouain,^{*} Mattia Pelizzola,[†] Monica Capozzoli,[†] Matteo Urbano,[†] Monique Capron,^{*} Filippo Belardelli,[‡] Francesca Granucci,[†] and Paola Ricciardi-Castagnoli^{3†}

- 2 experiments with design: Stimulation of DCs with *Schistosoma* + several time points + 2 biological replicas per time point
- Main results was:
 - A set of 283 DEGs
 - Functional classification of the first 98 DEGs
 - “Taken as a whole, our data provide molecular insights into the immune evasion mechanism of *schistosomula* and suggest an unexpected role for type I IFN in the innate response to **helminth eggs**”
 - **CausalAssertion, subject:**Helminthiasis(UMLS_ST:T047), **target:** IFN I response (GO:0032606)

An real case stored in MannMakna

makna Edit: SchistosomaPaper

http://localhost:8080/mannMakna/wiki/SchistosomaPaper?do=Edit

makna Edit: SchistosomaPaper

JSPWiki: Image

Find!

makna

Main page

About

SemWikiSyntax

Recent Changes

Unused pages

Undefined pages

Page Index

Help on editing

Referenced by
...nobody

makna v0.0.2 based on JSPWiki
2.2.28
[System info](#)
[Smallpiece Template](#)

Edit SchistosomaPaper

[mann#DocumentEntity!=]

!!!A Type I IFN-Dependent Pathway Induced by Schistosoma mansoni Eggs in Mouse Myeloid Dendritic Cells Generates an Inflammatory Signature.

[EGNPS1!mann#entityIsPositivelySupportedBy]

[[Image src='SchistosomaPaper/schisto_clusters.jpg' caption='Clusterization of 98 genes' width='700']]

Save Preview Cancel

An real case stored in MannMakna

makna: SchistosomaPaper

http://localhost:8080/mannMakna/wiki/SchistosomaPaper?useinference=true

makna: SchistosomaPaper

JSPWiki: Image

makna

Find!

Try a Semantic Search Add a statement about Schistosoma Delete a statement Edit this page More info Attach file

SchistosomaPaper

DocumentEntity type

A Type I IFN-Dependent Pathway Induced by Schistosoma mansoni Eggs in Mouse Myeloid Dendritic Cells Generates an Inflammatory Signature.

EGNPS1★

Properties of this page:

Predicate	Object
type	DocumentEntity
entityIsPositivelySupportedBy	EGNPS1
entityIsSupportedBy	EGNPS1
type	Entity
type	Resource
type	CommunityEntity
type	CommunityArtifact

Statements about this page:

no matching statements found. Why don't you add one?

Suggested actions:

[show only explicit](#)

Search for Resources with

- type [DocumentEntity](#)
[Explicit](#) [Inferenced](#)
- [entityIsPositivelySupportedBy](#) [EGNPS1](#)
[Explicit](#) [Inferenced](#)
- [entityIsSupportedBy](#) [EGNPS1](#)
[Explicit](#) [Inferenced](#)
- type [Entity](#)
[Explicit](#) [Inferenced](#)
- type [Resource](#)
[Explicit](#) [Inferenced](#)
- type [CommunityEntity](#)
[Explicit](#) [Inferenced](#)

makna v0.0.2 based on JSPWiki 2.2.28
[System info](#)

The Makna-Design is based on the [Smallpiece Template for JSPWiki](#)

Referenced by ...nobody

Time after parasite encounter (hrs)

Eggs SLA Cluster

Innate inflammatory response

Antigen presentation and costimulation

Cytoskeletal rearrangement and cell motility

Growth arrest and apoptosis

Additional immunologically relevant receptors

Iron response

Factors

92694_at CHI3I3 / YM-1 / ECF-L
101160_at CXCL2 / MIP-2
102629_at TNF-α
94146_at CCL4 / MIP-1β
103466_at IL-1β
95349_g_at CXCL1 / GRO-1
102424_at CCL3 / MIP-1α
103448_at MRP8 / Calgranulin A
93871_at IL-1ra
92849_at CCL8 / C10
95348_at CXCL1 / GRO-1
162198_f_at CCL8 / C10
160101_at Hmox1
104398_at CCL9 / MIP-1γ
93717_at CCL12 / MCP-5
93858_at CXCL10 / IP-10
94166_g_at CCL1 / I-309
100779_at IL-12p40
AFFX-MurIL2_at IL-2
94755_at IL-1α
102310_at CCL22 / ABCD-1
104443_at CCR7 / EBI-1
98406_at CCL5 / RANTES
99991_at IL-17r
95597_at Ptpn1 / COX-1
101918_at TGF-β1
163256_at LTβ4R1 / BLTR

93810_at Cathepsin D
160406_at Cathepsin K
103016_s_at CD68
101963_at Cathepsin L
93865_s_at H-2T10
101876_s_at H-2T17
101300_at H-2D2, Thy19.4
96752_at ICAM-1
92962_at CD40
104429_at H-2DOj

103855_at Plectin 1
97519_at Osteopontin
96747_at Arhu / Wrch1
102249_at Advillin
99799_at Vav oncogene
104681_at Diap1
98001_at Arhgef1
92280_at Actinin, α1

94928_at TNF-R2, p75 / CD120b
94186_at TRAF1
161666_f_at GADD45j
97504_at Cyclin D2
98067_at p21(WAF1/Cip1)

103335_at Lgals9 / galectin-9
95951_at Macrophage C-type lectin
102337_s_at FcγRIIb / CD32
102663_at uPAR / Ly-6 / CD87
98088_at CD14
102974_at MARCO
93454_at C1qr1 / Ly-68
99669_at Lgals1 / galectin-1
103005_s_at CD44
98828_at Mac-1 / CD11b

103990_at Fcεβ
92614_at Id3
92399_at Runx1 / CBFA2 / AML1
98875_at hr / hairless
93465_at EST, similar to: rat jdp-1
104104_at Spermatogenic Zip 1

Find: support

Next Previous Highlight all Match case Reached end of page, continued from top

Done


A Demo Application: MannMakna Wiki



Try a Semantic Search Add a statement about SchistosomaConclusions Delete a statement Edit this page More info Attach file

SchistosomaConclusions

GeneExpressionAssertion^{type}

The role of IFN/I in the innate response to helminth eggs

The following conclusion has been given in the [Schistosoma paper](#) .

taken as a whole, our data provide molecular insights into the immune evasion mechanism of schistosomula and suggest an unexpected role for [type I IFN](#)  in the innate response to [helminth eggs](#) .

The assertion is supported by [EGNPS1](#)  and [EGNPS2](#)  experiments.

[Go to top](#) [Edit this page](#) [More info...](#) [Attach file...](#)

This page last changed on 10-Nov-2007 18:54:46 GMT by 127.0.0.1.

Properties of this page:

Predicate	Object
seeAlso	Schistosoma
entityIsPositivelySupportedBy	EGNPS1
entityIsPositivelySupportedBy	EGNPS2
type	GeneExpress
assertionSubject	GO_003260
assertionContext	NCI_C35002

Statements about this page:

no matching statements found. Why don't you add one?

Suggested actions:

[show explicit+inferred](#)

Search for Resources with

- [seeAlso](#) [SchistosomaPaper](#)
[Explicit](#) [Inferred](#)
- [entityIsPositivelySupportedBy](#) [EGNPS1](#)
[Explicit](#) [Inferred](#)

A Demo Application: MannMakna Wiki

Firefox File Edit View History Bookmarks Tools Window Help

NetBeans IDE 5.5

makna: EGNPS1_tfs_Egg_12h_low

http://localhost:8080/mannMakna/wiki/EGNPS1_tfs_Egg_12h_low

makna: EGNPS1_tfs_Egg_12h_low x makna: EGNPS1_tfs_Egg_24h_null x

makna

Try a Semantic Search Add a statement about EGNPS1_tfs_Egg_12h_low Delete a statement Edit this page More info Attach file

EGNPS1_tfs_Egg_12h_low

[ExpressionFromTechnology](#)^{type} About the experiment: [E-GNPS-1](#)

D1 + Eggs or SLA (Functional classification of 98 differentially expressed genes)

16 genes encoding Additional transcription factors

Stimulus: [Schistosoma mansoni Eggs](#)★, **Time:** [12h](#)★, **Category:** [tfs](#)★

The genes have been computed as Under expressed
Average Expression level in this set: "0.643"★

Support:

These set has been computed from the raw data files: Experiment E-GNPS-1 [12h, Egg, rep. 1](#)★
Experiment E-GNPS-1 [12h, Egg, rep. 2](#)★

Genes:

Egr-2, Zfp-6, Krox20, Zfp-25, Krox-20 ★	0.78
hr ★	0.76
Fosb ★	0.73
Idb3 ★	0.71
AML1, CBFA2, AMLCR1, PEBP2A2, PEBP2aB ★	0.64
RIKEN cDNA 9130211I03 gene ★	0.51
Spermatogenic Zip 1 ★	0.37

Properties of this page:

Predicate	Object
assertionSubject	_at
expressionData	E-GNPS-1-H7.cel
assertionSubject	_at
assertionSubject	_at
assertionSubject	_at
assertionContext	Time12h
assertionContext	Eggs
type	ExpressionFromTechnology
expressionData	E-GNPS-1-H8.cel
intensity	0.643
assertionSubject	_at
assertionContext	tfs
assertionSubject	_at
assertionSubject	_at

Statements about this page:

no matching statements found. Why don't you add one?

Suggested actions:

[show explicit+inferred](#)

Search for Resources with

Go to top Edit this page More info... Attach file...

Find: broader Next Previous Highlight all Match case

makna v0.0.2 based on JSPWiki 2.2.28
[System info](#)

The Makna-Design is based on the [Smallpiece Template for JSPWiki](#)

A Demo Application: MannMakna Wiki

Firefox File Edit View History Bookmarks Tools Window Help

NetBeans IDE 5.5
makna: 103990_at

http://localhost:8080/mannMakna/wiki/103990_at

makna: 103990_at x makna: EGNPS1_tfs_Egg_24h_null x

Google

makna

Try a Semantic Search Add a statement about 103990_at Delete a statement Edit this page More info Attach file

103990_at

Main page
About
SemWikiSyntax
Recent Changes
Unused pages
Undefined pages
Page Index
Edit this page
Set your name

Referenced by
...nobody

makna v0.0.2 based on JSPWiki 2.2.28
[System info](#)

The Makna-Design is based on the [Smallpiece Template for JSPWiki](#)

Properties of this page:

Predicate	Object
entityTermAnnotation	GO_0003750
entityTermAnnotation	GO_0008151
entityTermAnnotation	GO_0005634
probedBy	U74Av2
type	ProbeSetContainer
entityTermAnnotation	GO_0005622
entityTitle	Fosb
entityTermAnnotation	GO_0006355
entityTermAnnotation	GO_0003700
entityName	103990_at
entityTermAnnotation	GO_0003677
entityTermAnnotation	MusMus

Statements about this page:

Subject	Predicate
EGNPS2_tfs_SLA_24h_null	assertionSubject
tfs_SLA_24h_null	assertionSubject
tfs_SLA_8h_high	assertionSubject
EGNPS2_tfs_SLA_8h_high	assertionSubject
EGNPS2_tfs_SLA_0h_high	assertionSubject
EGNPS2_tfs_SLA_4h_null	assertionSubject
tfs_SLA_12h_null	assertionSubject

103990_at

ProbeSetContainer^{type}

"103990_at"[★]

"Fosb"[★]

[U74Av2](#)[★]

[MusMus](#)[★] (this may be inferred by the chip)

GO Biological Process:

[transcription regulation](#)[★]

[cell growth and/or maintenance](#)[★]

GO Cellular Component

[nucleus](#)[★]

[intracellular](#)[★]

GO Molecular Function

[transcription factor](#)[★]

[DNA binding](#)[★]

[cell cycle regulator](#)[★]

Go to top Edit this page More info... Attach file...

This page last changed on 16-Jul-2007 12:53:21 BST by 0:0:0:0:0:0:1%0.

X Find: broader Next Previous Highlight all Match case

Done

Prepared queries: genes -> conditions

```

SELECT
  DISTINCT ?pbset ?name ?title ?geAss ?ctx ?ctxTerm
WHERE
{ ?pbset rdf:type mann:ProbeSetContainer
  .
  { # _____ Basic Attrs _____
    { ?pbset mann:entityTitle ?title . FILTER regex( ?title, "$keywords", "i" )
      . OPTIONAL { ?pbset mann:entityName ?name }
    } UNION {
      ?pbset mann:entityName ?name . FILTER regex( ?name, "$keywords", "i" )
      . OPTIONAL { ?pbset mann:entityTitle ?title }
    }
  }
  # _____ Term Annotations _____
  UNION {
    ?pbset mann:entityTermAnnotation ?term
    . ?term mann:entityTitle ?termTitle
    . FILTER regex( ?termTitle, "$keywords", "i" )
    . OPTIONAL { ?pbset mann:entityTitle ?title }
    . OPTIONAL { ?pbset mann:entityName ?name }
  }
}

# _____ Where they are expressed _____
.
OPTIONAL {
  ?geAss rdf:type mann:Assertion;
  mann:assertionSubject ?pbset
  . OPTIONAL {
    ?geAss mann:assertionContext ?ctx . ?ctx
    mann:geExperimentalFactorAnnotation ?ctxTerm
  }
}
}

```

- Parameters coming from user Input (simple code written)
- Works over the inferred model
- Results may be integrated into the Wiki (search function and current page statements)

Prepared queries: conditions->genes

```

SELECT
  DISTINCT ?pbset ?pbsTitle ?geAss ?ctx ?ctxTerm ?ctxTermTitle ?expLevel
WHERE
{
  ?geAss rdf:type mann:GeneExpressionAssertion;
    mann:assertionSubject ?pbset . ?pbset rdf:type mann:ProbeSetContainer

  . OPTIONAL { ?pbset mann:entityTitle ?pbsTitle }

  . OPTIONAL {
    ?geAss mann:intensity ?expLevel
    . FILTER ( xsd:float ( ?expLevel ) > $level )
  }

  .
  {
    { # ____ Match the URI ____
      ?geAss mann:assertionContext ?ctx
      . ?ctx mann:geExperimentalFactorAnnotation ?ctxTerm
      . FILTER regex ( str ( ?ctxTerm ) , "$keywords", "i" )
      . OPTIONAL { ?ctxTerm mann:entityTitle ?ctxTermTitle }
    }

    UNION
    { ____ Or match the Title ____
      ?geAss mann:assertionContext ?ctx
      . ?ctx mann:geExperimentalFactorAnnotation ?ctxTerm
      . ?ctxTerm mann:entityTitle ?ctxTermTitle . FILTER regex( ?ctxTermTitle,
"$keywords", "i" )
    }
  }
}

ORDER BY DESC(xsd:float(?expLevel))

```

Prepared queries

Find Genes expressed under given conditions – MannMakna

http://localhost:8080/mannMaknaExtra/query/condGenes/show.jsp?keyword=Inflammation&level= Google

Find Genes expressed under given conditions, Results:

[mkb:92694_at](#), Chitinase 3-like 3

Involved/Expressed in: [mkb:EGNPS1_inflammation_Egg_0h_high](#), context: [mkb:Inflammation](#) condition: [mkb:InflammationTerm](#), level: 1.74

[mkb:95597_at](#), PGG/H

Involved/Expressed in: [mkb:EGNPS1_inflammation_Egg_0h_high](#), context: [mkb:Inflammation](#) condition: [mkb:InflammationTerm](#), level: 1.74

[mkb:99991_at](#), Il17r

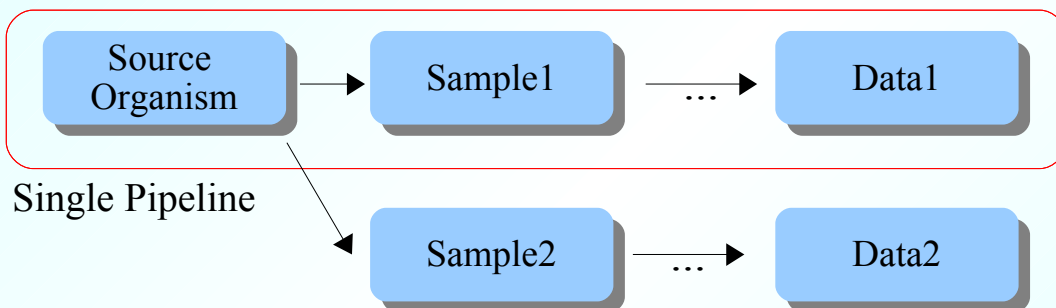
Involved/Expressed in: [mkb:EGNPS1_inflammation_Egg_0h_high](#), context: [mkb:Inflammation](#) condition: [mkb:InflammationTerm](#), level: 1.74

Done

MANN Rules: Reified relations

```
-> (mann:NormalizedData mann:usesMaterialInPipelines mann:RawData).
-> (mann:HybridizationData mann:usesMaterialInPipelines mann:BiologicalMaterial).
-> (mann:HybridizationData mann:usesMaterialInPipelines mann:LabeledExtract).
-> (mann:LabeledExtract mann:usesMaterialInPipelines mann:Extract).
...

#
# Inference about Material usage:
# If x type Class1, y type Class2, Class1 usesMaterialInPipelines Class2,
#   x,y in the same pipeline
# THEN x uses y
#
# For instance: a sample uses a source that is in the same pipeline
#
[materialPipelineUse:
  (?matx mann:usesGEntity ?maty)
  <-
    (?ep mann:pipelineMaterial ?matx),
    (?ep mann:pipelineMaterial ?maty),
    (?matx rdf:type ?MatTypeX),
    (?maty rdf:type ?MatTypeY),
    (?MatTypeX mann:usesMaterialInPipelines
      ?MatTypeY)
]
```



Try a Semantic Search Add a statement about EGNPS2_Pipeline4

EGNPS2_Pipeline4

Experiment^{type} for the Experiment **E-GNPS-2**

Source: [SRC-GNPS-E2-1](#)★
 Grow Protocol: [P-GNPS-GRWCND-E2-SRC1](#)★
 Sample: [SMP-GNPS-E2-2](#)★
 Treatment Protocol: [P-GNPS-TRT-E2-SRC1-SMP2](#)★
 Extraction Protocol: [P-GNPS-EXT2](#)★
 Extract: [XTR-GNPS-E2-2](#)★
 Labeled Extract: [LBL-GNPS-E2-2](#)★
 Hybridization Protocol: [P-GNPS-HYB6](#)★
 Hybridization: [HYB-GNPS-E2-3](#)★
 Scanning Protocol: [P-GNPS-SCN7](#)★
 Image Analysis Protocol: [P-GNPS-IMG8](#)★
 Raw Data: [EGNPS2H3.cel](#)★
 Normalization Protocol: [P-GNPS-NRM9](#)★
 Final (Normalized) Data: [E-GNPS-2_fdm.txt](#)★

[Go to top](#) [Edit this page](#) [More info...](#) [Attach file...](#)

This page last changed on 16-Aug-2007 23:00:35 BST by 127.0.0.1.

Microarrays and Gene Expression Analysis

Formal models in Microarray Knowledge

Our Semantic Web based proposal

A demo application

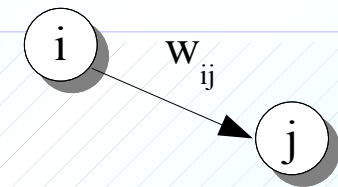
A proposal for Knowledge Ranking

Conclusions/Future

Spread Activation Search/Ranking

Basic SA

$$I_j = \sum_i w_{ij} \cdot O_i \cdot \beta$$



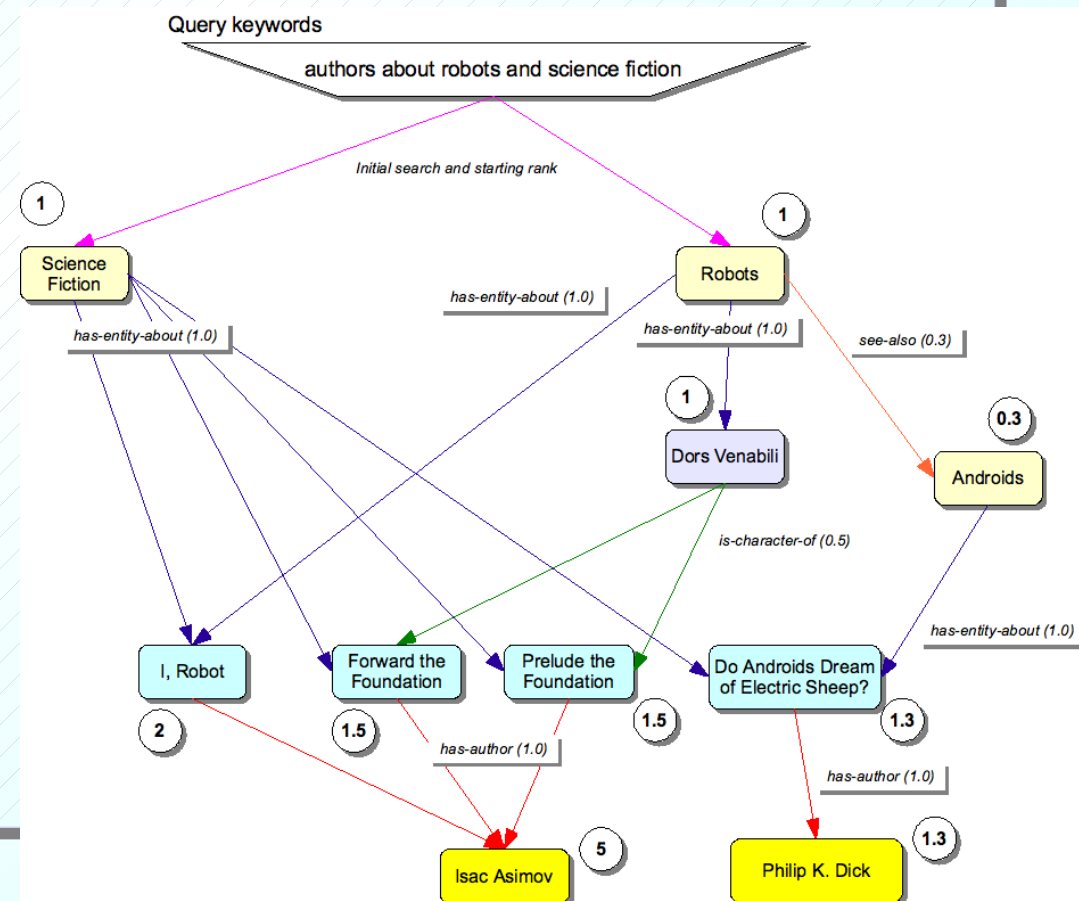
The ranking of node j is a weighted sum of the ranking of incoming nodes. The weight depends on the edges, for instance on the relation type.

$I_i = f(O_i)$ in the simplest case $I_i = O_i$

```

queue = initialQueue(); stopFlag = false;
while ( !queue.empty() && !stopFlag )
    i = queue.pull()
    if ( checkPreRestrictions(i) )
        for each j in (i,j)
            j.in += w(i,j) * i.out * beta
            j.out = f(j.in)
            if ( !j.visited )
                queue.push ( j )
        }
    }
stopFlag = checkPostRestrictions()
}

```



Spread Activation Search/Ranking

- May be generalised to RDF graphs, with an available Query language (SPARQL)
- We may define a set of SPARQL queries, which receive the current node i as parameter and return a set of semantically linked “outgoing” nodes j : $\{Query_k(i)\}$
- The propagation of activation in the algorithm is done according to:

$$\forall j \in Query_k(i): I'_j = I_j + w_k \cdot O_i \cdot \beta$$

- Initial ranking too may be arranged with SPARQL or rules
- Again, it may well work with the inferred graph
- May be flexibly used for either ranking or searching from an initial set of nodes

SW Spread Activation

• Initialisation, evaluated entities

```
For ech x in:
  SELECT ?x WHERE ?x mann:evaluation ?v
do
  x.out += v
```

• Initialisation, evaluated entities

```
For each x in:
  SELECT ?x WHERE ?x rdf:type mann:Assertion
  x.out += 1
```

• may use inference (sub-classes of Assertion)

• Propagation of evaluations given by comments

```
Starting from c, for each x in:
  SELECT ?x WHERE
    $c rdf:type mann:Comment
    $c mann:assertionSubject ?x
do
  x.out += 0.8 * c.out * beta
```

• Propagation of support

```
Starting from s, for each x in:
  SELECT ?x WHERE $s mann:entityPositivelySupports ?x
do
  x.out += s.out * 0.8 * beta
```

• may use inference (sub-properties)

• similar query for negative support (neg. weight)

SW Spread Activation

• Author-based ranking

```
Starting from a, for each x in:  
  SELECT ?x WHERE  
    $a rdf:type mann:Person  
    ?x mann:Experiment  
    ?x mann:hasPrincipalInvestigator $a  
do  
  x.out += $a.out * beta
```

- similar for hasInvestigator (with minor weight)
- may be refined considering Person (student, professor, etc.) or the role played by the person (worksWith etc.)
- Other approaches may be used to provide an initial ranking of persons or their publications (e.g.: co-citation, IF, Social Network Analysis)

SW Spread Activation

- A rule for gene assertion counting

```
[ (?ass mann:assertionSubject ?pbset)
  (?ass mann:assertionContext ?ctx)
  (?ctx mann:entityTermAnnotation ?term)
=>
  addSupportToGExpression ( ?pbset, ?term, ?ass )
```

- For each new pair of probeset/term, adds/create the statements:

```
[<id> rdf:type InferredGExpression
  mann:assertionSubject ?pbset
  mann:assertionContext [ <ctx:id> mann:entityTermAnnotation ?term]
  mann:supportedBy ?ass]
```

- i.e.: sum-up pairs of probeset/term
- the propagation of support will automatically weight this synthesising assertion, the more it is asserted the more it is ranked
- if the original assertion is ranked according to several criteria (e.g.: author, evaluation, etc.), then this is propagated too

Microarrays and Gene Expression Analysis

Formal models in Microarray Knowledge

Our Semantic Web based proposal

A demo application

A proposal for Knowledge Ranking

Conclusions/Future

Some words of conclusion

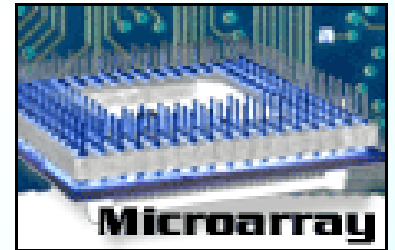
- **The Semantic Web is increasingly being used in Life Sciences**
 - mostly because of heterogeneous information to be integrated
- **Ontologies are common as well, reasoning will hopefully be used** more in the future of Life Sciences
 - Expressive formalization of Biological Knowledge
 - Even the basic inference is useful
 - Problems to be solved: scalability, performance
- **We have shown an example** of all of the above
 - A simple OWL model that reuse existing ontologies and formalise a piece of knowledge previously not machine-readable
 - A simple demo application that used the model
 - Examples of how to exploit the semantic content provided by the model

Possible future developments

- **Refine** MANN and integrate existing ontologies
- Try to reduce verbosity (of assertions)
- More Ontologies Integration
- Experiment with **SW/SA**
- From the demo to a more useable application
- **Better import** services and **Integration** in other Microarray Management Systems (e.g. BASE)
- Improve the UI, **AJAX** and **Visual editors**

Thanks to...

Ugis Sarkans
Alvis Brazma
all people by the Microarray Group @ EBI



Paola Ricciardi - Castagnoli
Andrea Splendiani
all people from the Castagnoli's group



Giancarlo Mauri, PhD supervisor
Mauro Pezzè, PhD mentor



Karsten Dello (Makna's main developer)