LSDIS

Large Scale Distributed Information Systems

University of Georgia
Computer Science Department

Semantic empowerment of Health Care and Life Science Applications

WWW 2006 W3C Track, May 26 2006

Amit Sheth

LSDIS Lab **&** **Semagix**

University of Georgia

Joint work with Athens Heart Center, and CCRC, UGA

Part I: A Healthcare Application

Active Semantic Electronic Medical Record @ Athens Heart Center (use Firefox)

(deployed since Dec 2005)

Collaboration between LSDIS & Athens Heart Center
(Dr. Agrawal, Dr. Wingate
For on line demo: Google: Active Semantic Documents

Active Semantic Document

A document (typically in XML) with

- Lexical and Semantic annotations (tied to ontologies)
- Active/Actionable information (rules over semantic annotations)

Application: Active Semantic EMR for Cardiology Practice

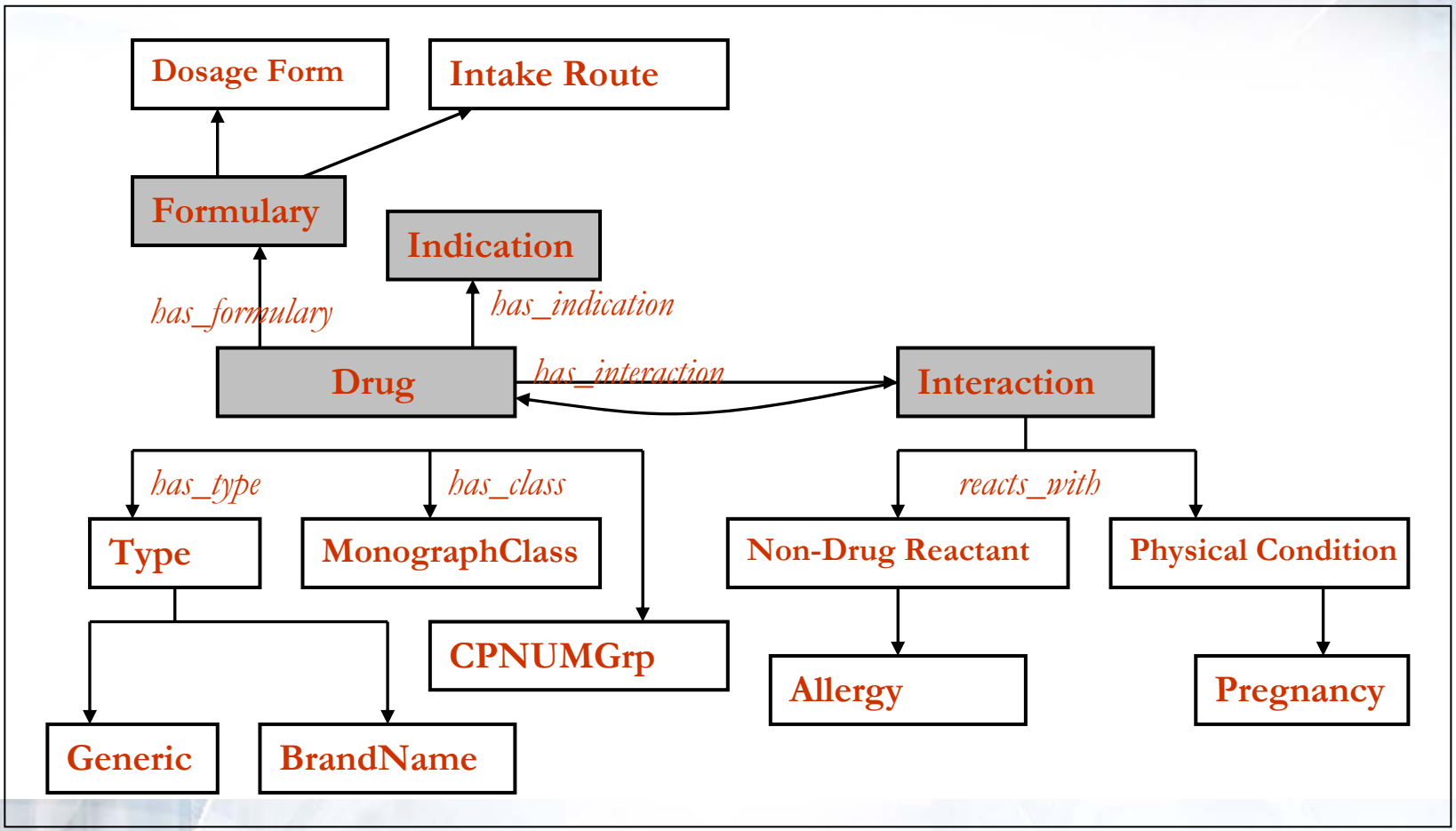
- EMRs in XML
- 3 ontologies (OWL), populated
- RDQL->SPARQL, Rules
- Services, Web 2.0

Active Semantic Electronic Medical Record

Demonstrates use of Semantic Web technologies to

- reduce medical errors and patient safety
  - accurate completion of patient charts (by checking drug interactions and allergy, coding of impression,...)
- improve physician efficiency, extreme user friendliness, decision support
  - single window for all work; template driven sentences, auto-complete, contextual info., exploration
- improve patient satisfaction in medical practice
  - Formulary check
- improve billing due to more accurate coding and adherence to medical guidelines
  - Prevent errors and incomplete information that insurance can use to withhold payment

One of 3 ontologies used (part of drug ontology)

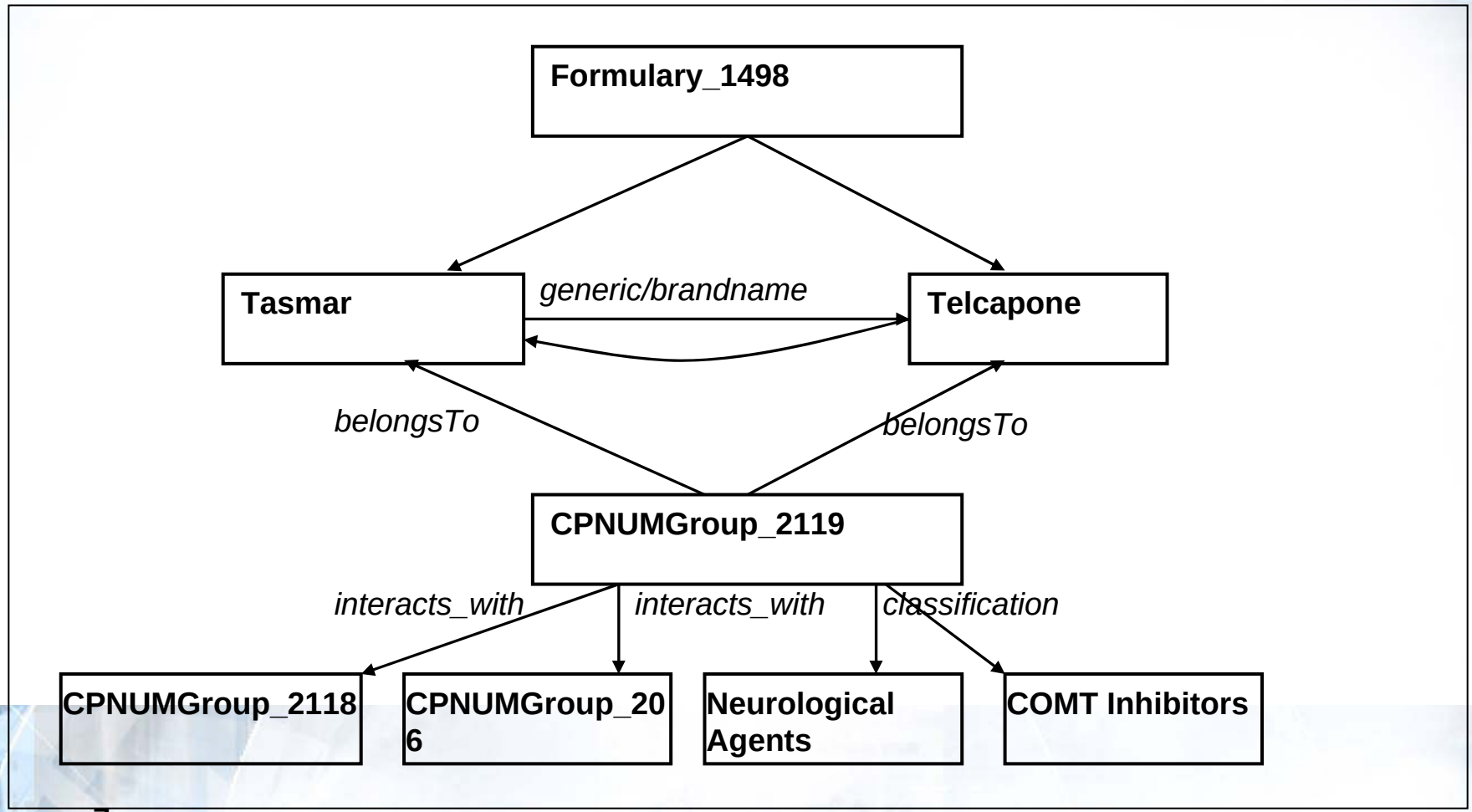


Local, licensed and public (Snomed) sources to populated ontologies

Example Rules

- drug-drug interaction check,
- drug formulary check (e.g., whether the drug is covered by the insurance company of the patient, and if not what the alternative drugs in the same class of drug are),
- drug dosage range check,
- drug-allergy interaction check,
- ICD-9 annotations choice for the physician to validate and choose the best possible code for the treatment type, and
- preferred drug recommendation based on drug and patient insurance information

Exploration of the neighborhood of the drug Tasmar



Active Semantic Doc with 3 Ontologies

Athens Heart Center 2005 Prince Avenue, Athens, GA, 30606 Phone: 706-208-9700 Fax: 706-208-0806

Jerek Chicken **Visit on 10/28/2005**

325 PileORice Street, Athens, GA 30606

SSN: 123-45-6789 **MR #: 555555** **Sex: M** **DOB: 01/02/1934** **Age: 71**

Other Physicians: Harry Wingate,M.D. E Kevin Adams,M.D. E
Family Practice
706-795-9588

Annotate ICD9s

Annotate Doctors

Problem List:

1. Hypertension **(365.04)** E
2. Cholecystectomy **(576.0)** E
3. Chest Pain E

Lexical Annotation

Chief Complaint: Evaluation of abnormal EKG status post abnormal Echo.Evaluation of aortic stenosis status post arterial examination.Cardiac clearance for aneurysm removal.Follow up of recent hospitalization at Barrow Community Hospital for acute myocardial infarction.

History of Present Illness: He was evaluated at Athens Regional Medical Center emergency room by Dr. Harry Wingate.He is here today for cardiac clearance for aneurysm removal.The patient reports chronic moderate burning and cramping chest pain located across the chest, which radiates to the arms. He reports that his chest pain is aggravated by movement. Th[illegible] breathing deeply.Patient's history is positive for the following cardiovascular risk factors: diabetes and family history of cor

Level 3 Drug Interaction

Insurance Formulary

Current Medications	Medications Af
Actos 30 mg,1tab E	Coumadin tablets 10 mg,1tab I3 F E Viagra 50 mg,1tab I3 F E Zyrtec 5 mg,1tab E Zyvox 2 mg/ml,1inj A E

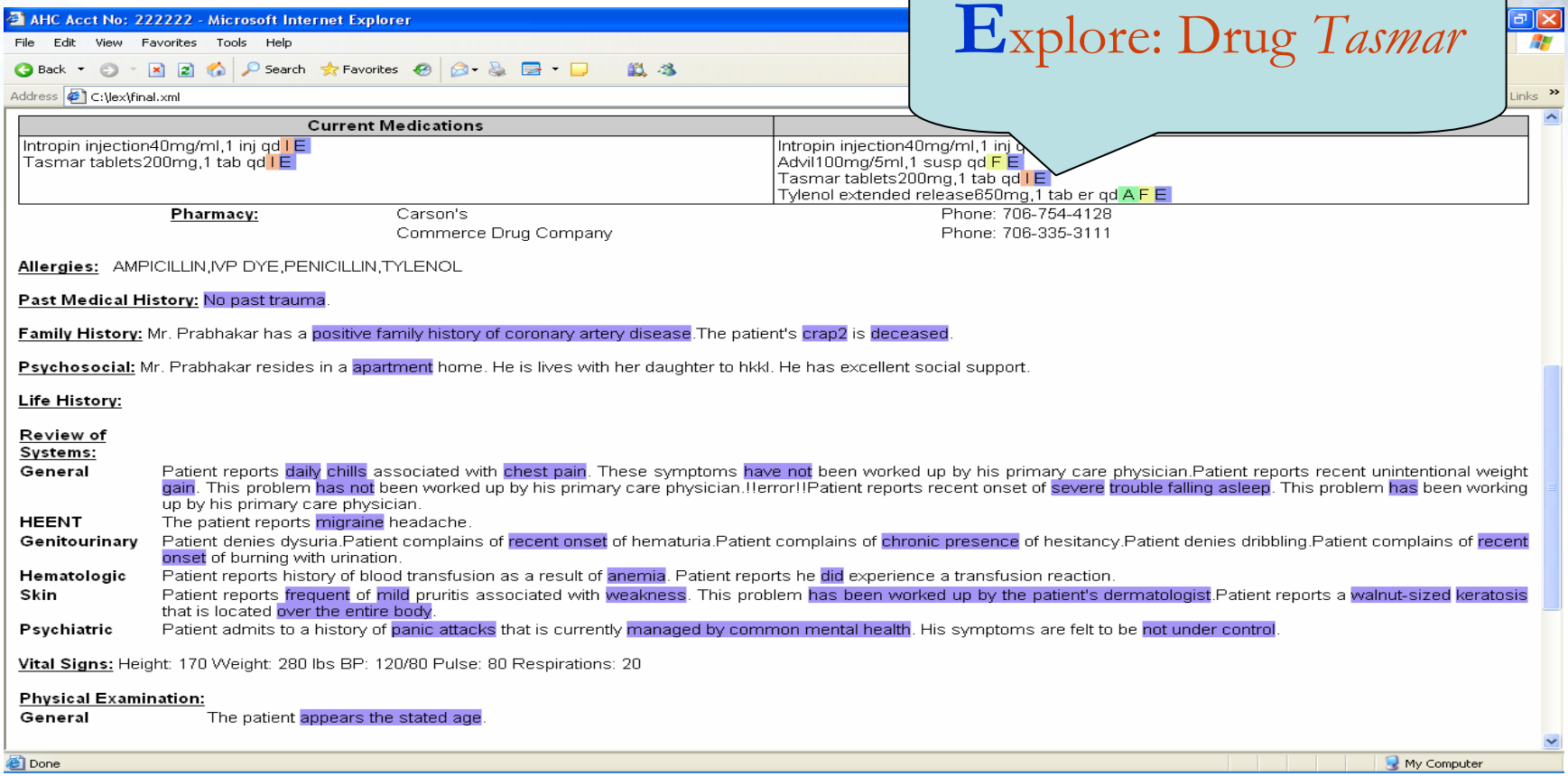
DrugAllergy

Allergies: LINEZOLID

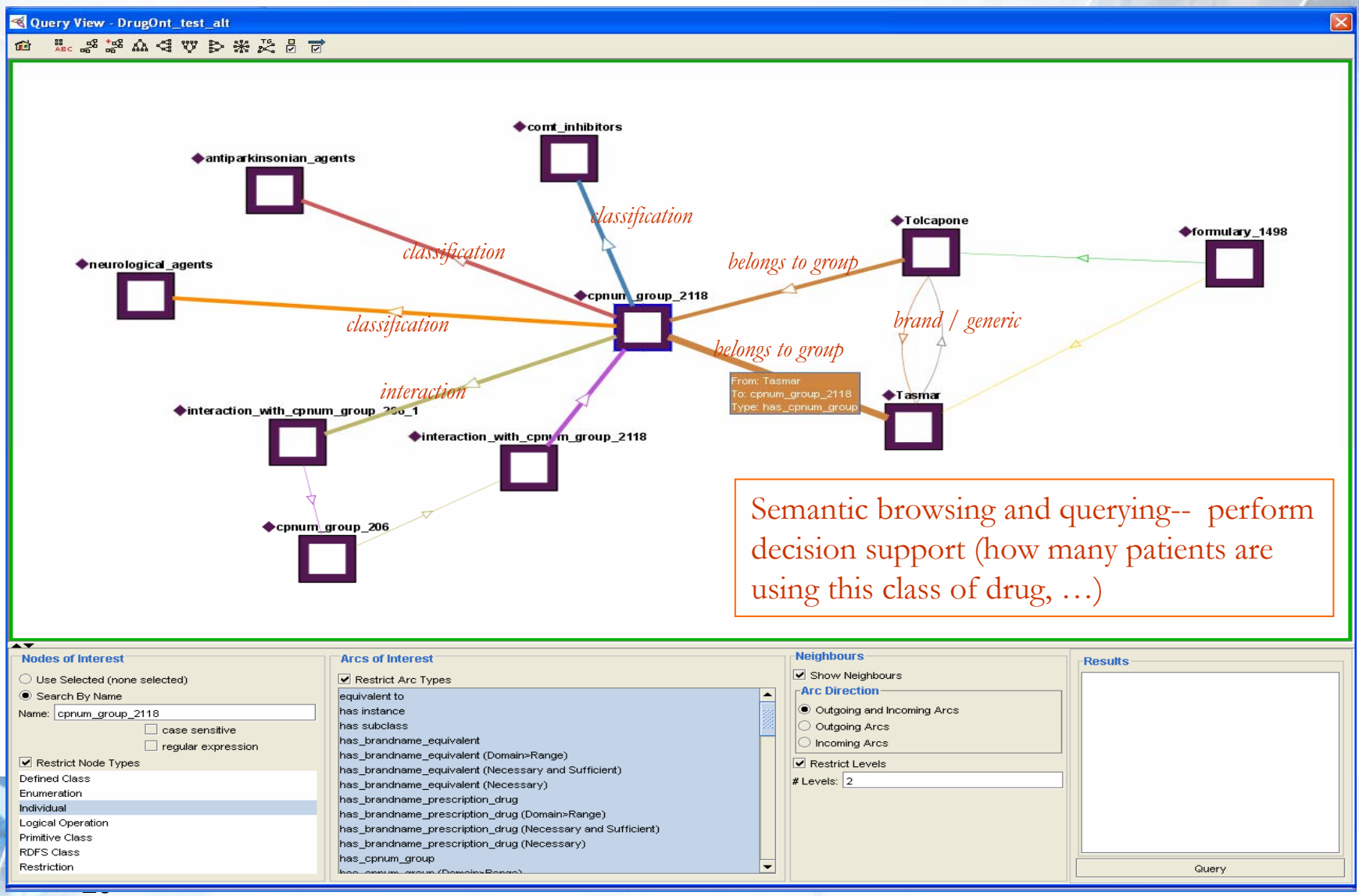
Impressions:

1. Abdominal aortic aneurysm, advanced secondary to by a positive nuclear scan.
2. Abnormal cardiac study associated with chest tightness appears to be secondary to a noncardiac cause as evidenced by arterial scan of lower extremities.
3. Normal cardiac cath

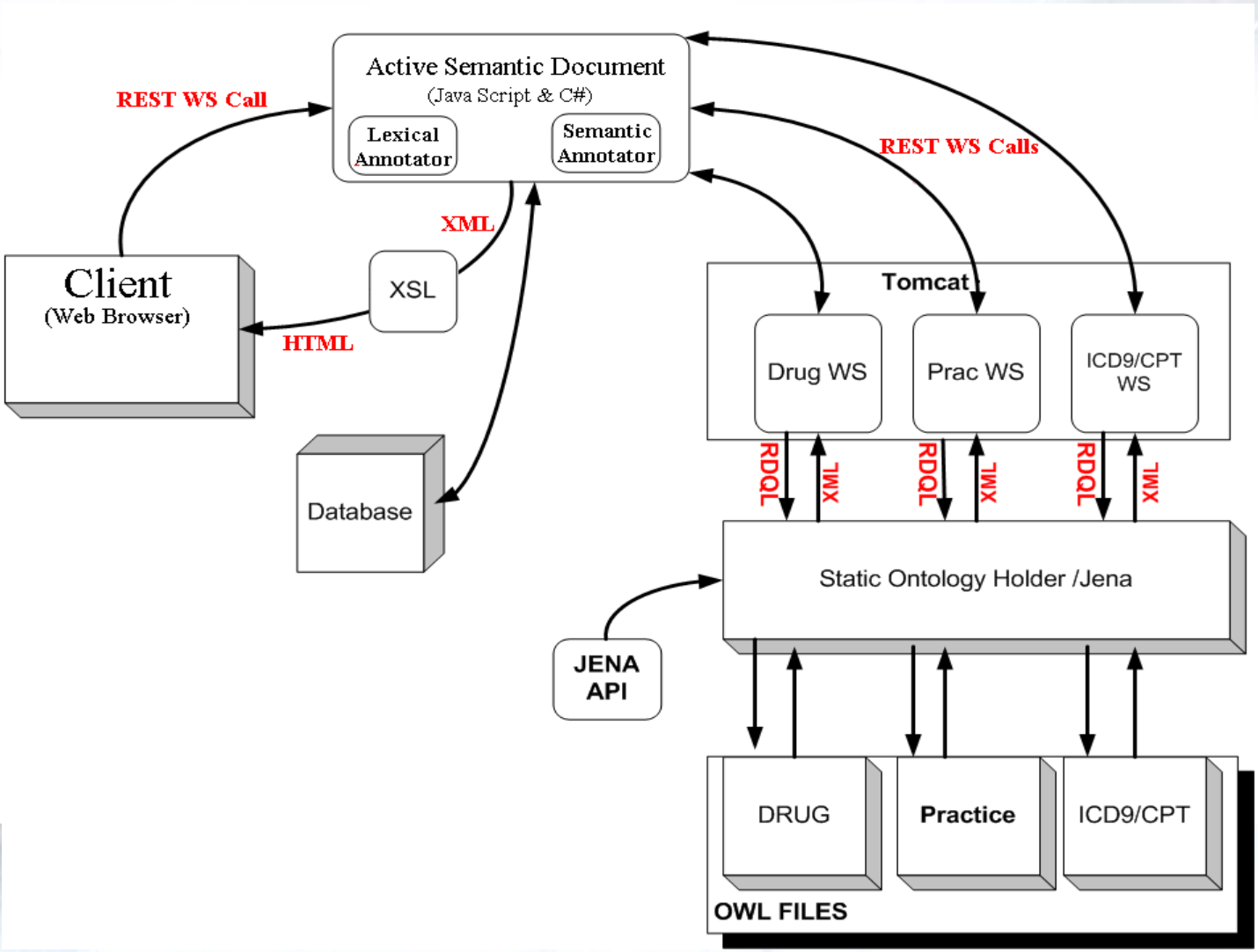
Explore neighborhood for drug *Tasmar*



Explore neighborhood for drug *Tasmar*

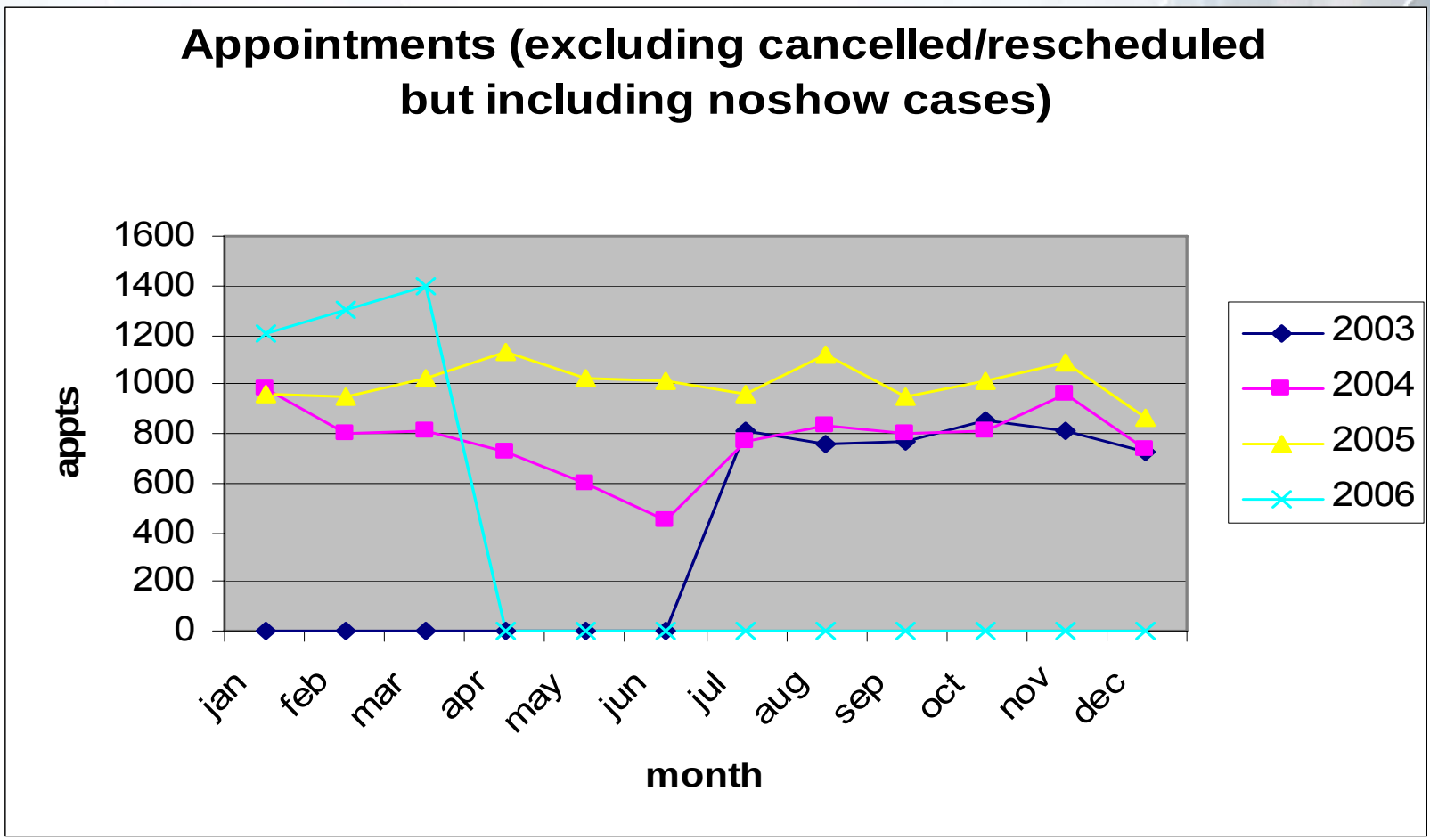


Software Architecture



ROI

Athens Heart Center Practice Growth



Increased efficiency demonstrated as more encounters supported without increasing clinical staff

Chart Completion before the preliminary deployment of the ASMER

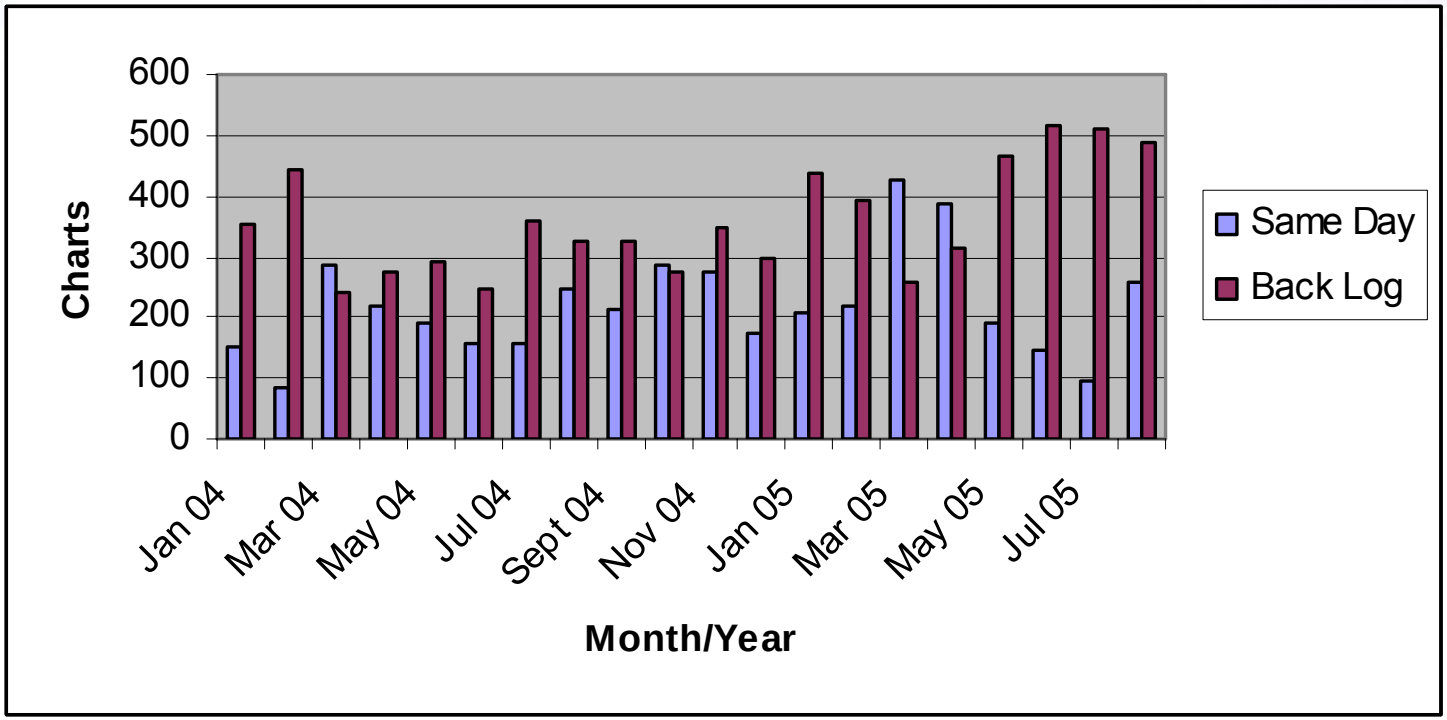
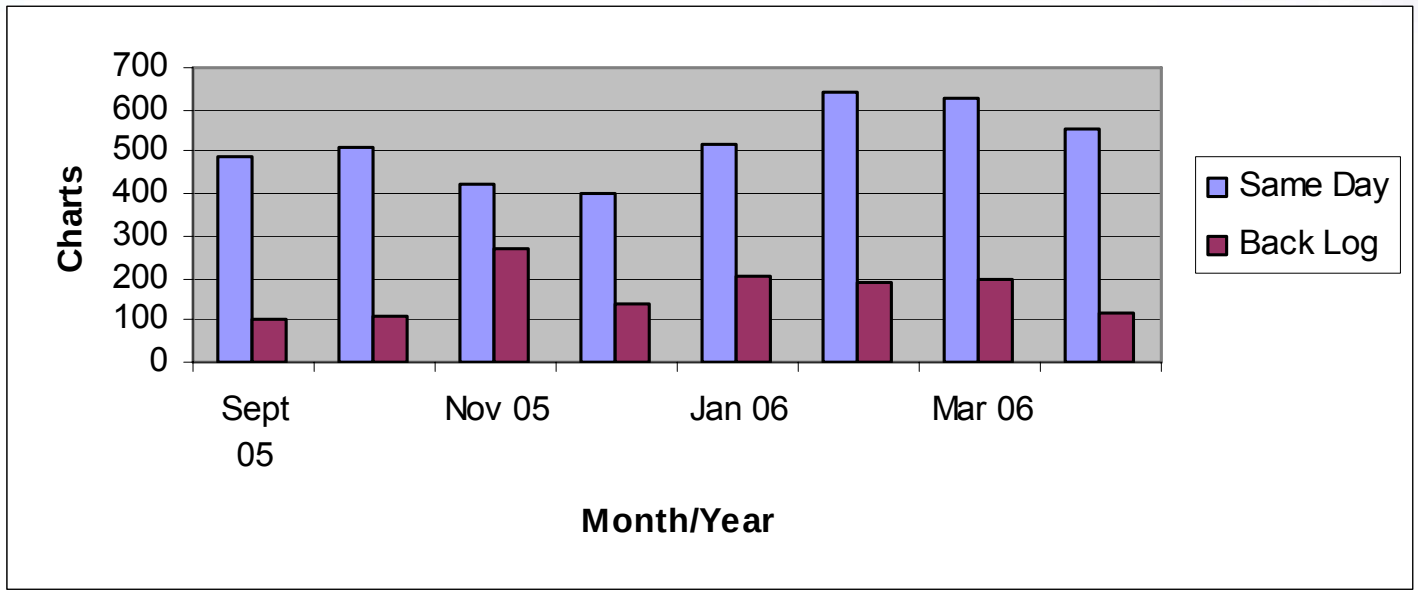
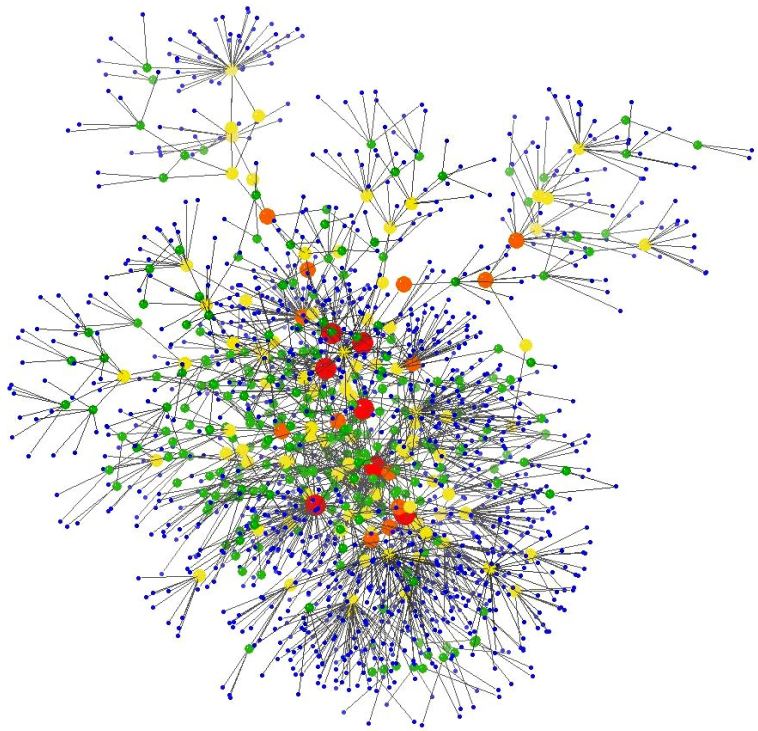


Chart Completion after the preliminary deployment of the ASMER

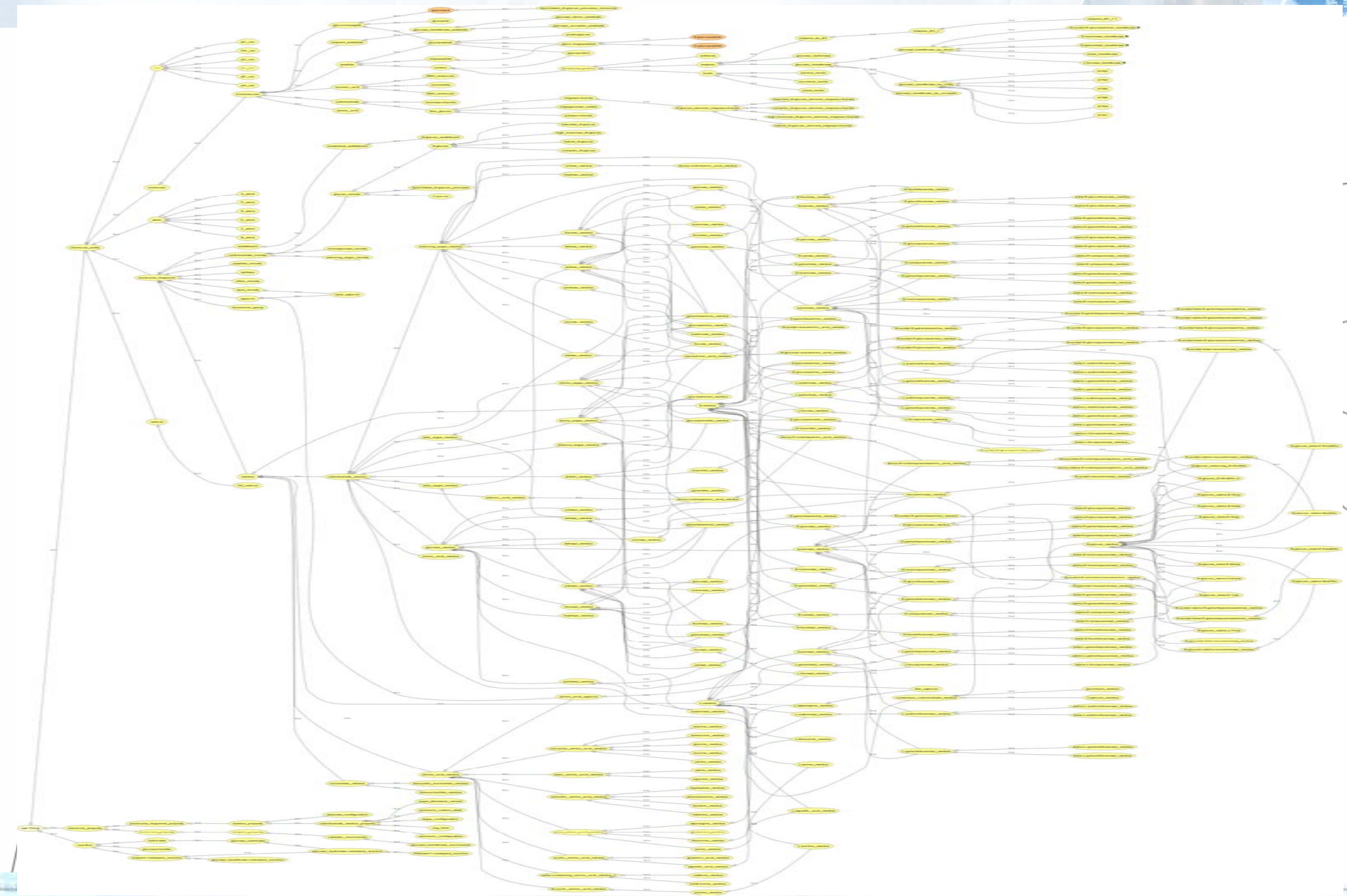


Applying Semantic Technologies to the Glycoproteomics Domain



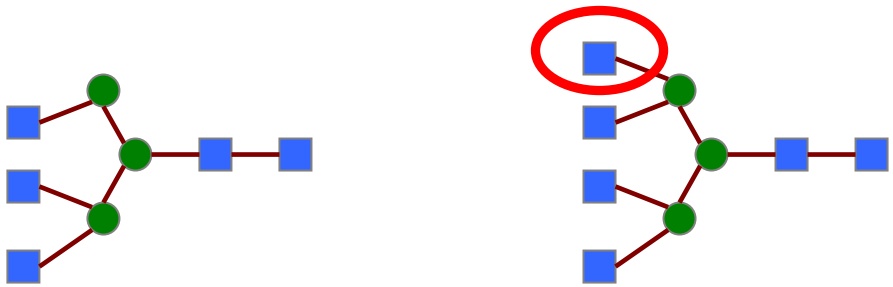
Quick take on bioinformatics ontologies and their use

- GlycO and ProPreO - among the largest populated ontologies in life sciences
- Interesting aspects of structuring and populating these ontologies, and their use
- GlycO
  - a comprehensive domain ontology; it uses simple 'canonical' entities to build complex structures thereby avoids redundancy → ensures maintainability and scalability
  - Web process for entity disambiguation and common representational format → populated ontology from disparate data sources
  - Ability to display biological pathways
- ProPreO is a comprehensive ontology for data and process provenance in glycoproteomics
- Use in annotating experimental data, high throughput workflow



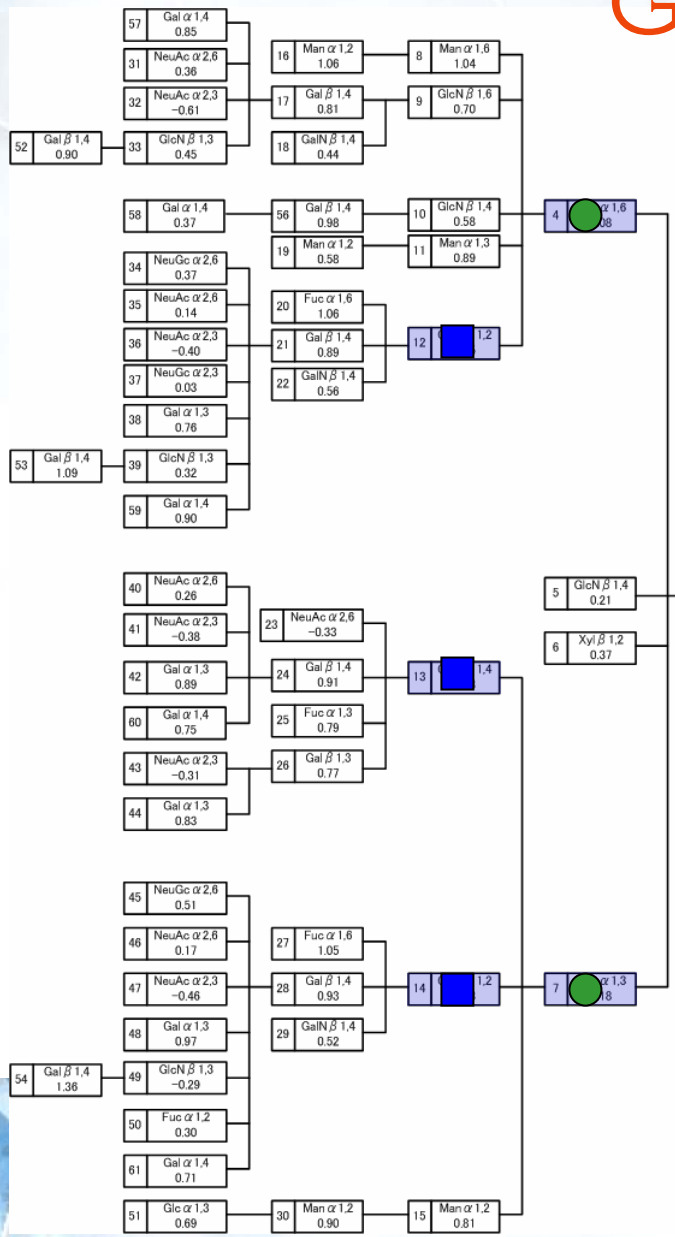
GlycO ontology

- Challenge – model hundreds of thousands of complex carbohydrate entities
- But, the differences between the entities are small (E.g. just one component)

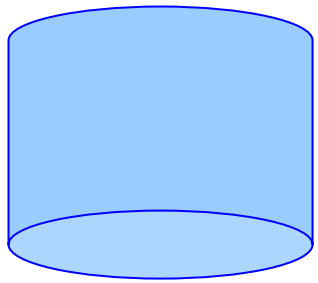


- How to model all the concepts but preclude redundancy → ensure maintainability, scalability

GlycoTree



β-D-Glc*p*NAc-(1-2)-α-D-Man*p*-(1-6)+
β-D-Man*p*-(1-4)-β-D-Glc*p*NAc-(1-4)-β-D-Glc*p*NAc
β-D-Glc*p*NAc-(1-4)-α-D-Man*p*-(1-3)+
β-D-Glc*p*NAc-(1-2)+

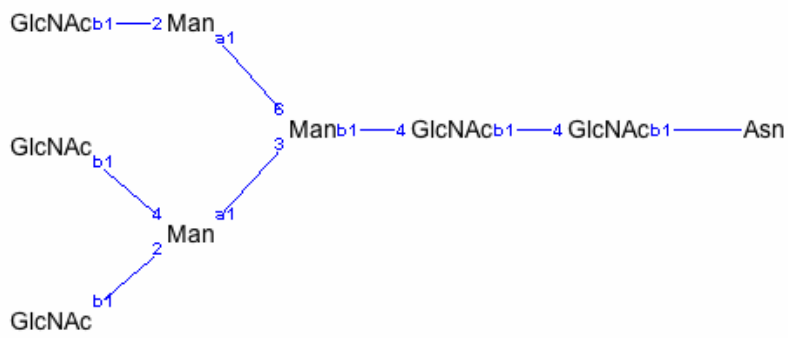


N. Takahashi and K. Kato, *Trends in Glycosciences and Glycotechnology*, 15: 235-251

Ontology population workflow

KEGG GLYCAN: G00020

Help

Entry	G00020 Glycan
Composition	(GlcNAc)5 (Man)3 (Asn)1
Mass	1502.4 (Asn)
Structure	 KCF file KCaM
Class	Glycoprotein; N-Glycan
Reaction	R05987 R05991
Pathway	PATH: map00510 N-Glycan biosynthesis
Enzyme	2.4.1.145 2.4.1.155
Ortholog	KO: K00738 alpha-1,3-mannosylglycoprotein beta-1,4-N-acetylglucosaminyltransferase KO: K00744 alpha-1,3(6)-mannosylglycoprotein beta-1,6-N-acetyl-glucosaminyltransferase
Other DBs	CCSD: 4362 5058 7636 30028 33048 34937 35723 42024 43245
LinkDB	All DBs
KCF data	Show

C to
UCS

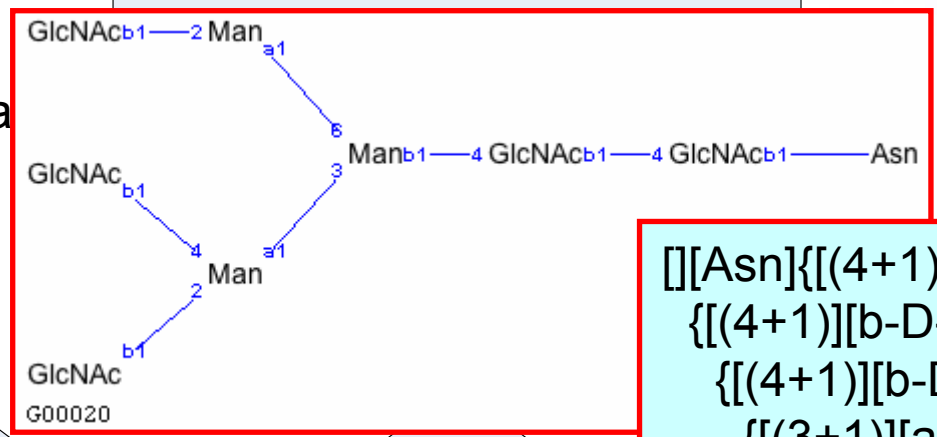
Base

CS to
GLYDE

GlycO population

Semagix Freedom knowledge extractor

YES:
next Insta



[][Asn]{[(4+1)][b-D-GlcpNAc]
{[(4+1)][b-D-GlcpNAc]
{[(4+1)][b-D-Manp]
{[(3+1)][a-D-Manp]
{[(2+1)][b-D-GlcpNAc]
{}[(4+1)][b-D-GlcpNAc]
{}}[(6+1)][a-D-Manp]
{[(2+1)][b-D-GlcpNAc]{}}}}}

Already in KB?

NO

Insert into KB

Has CarbBank ID?

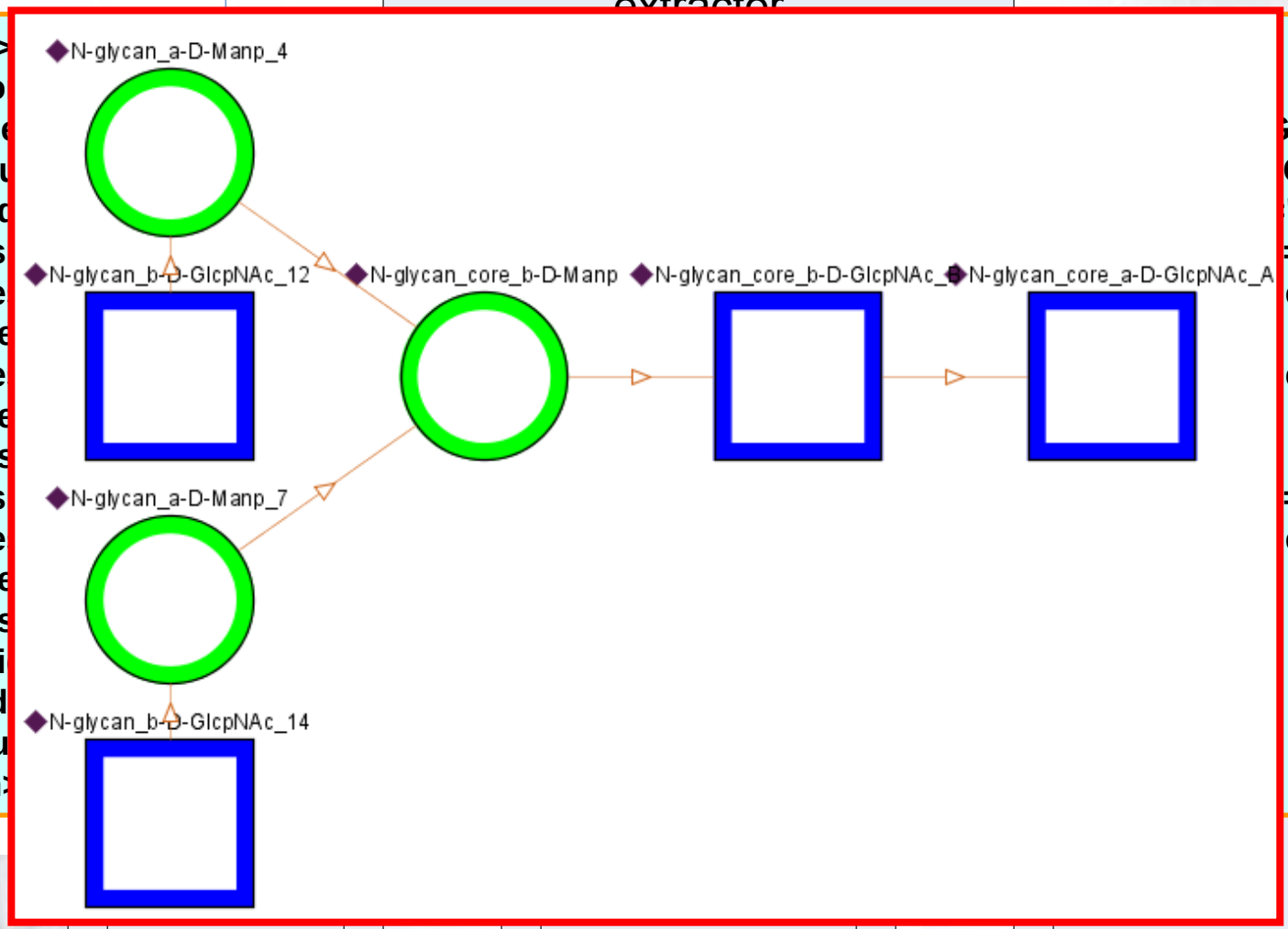
YES

Compare to Knowledge Base

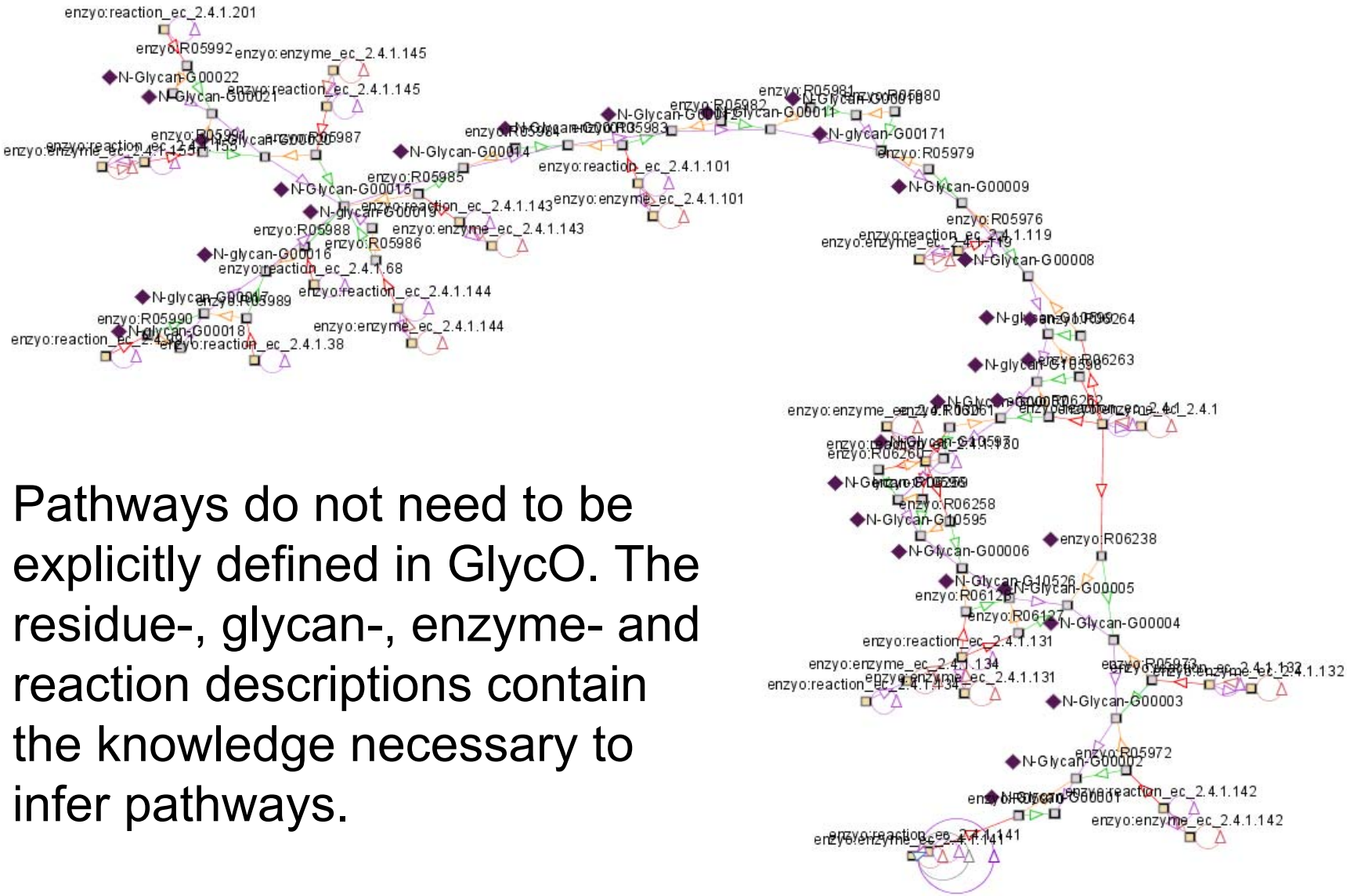
LINUCS to GLYDE

Ontology Population Workflow

Semagix Freedom knowledge extractor

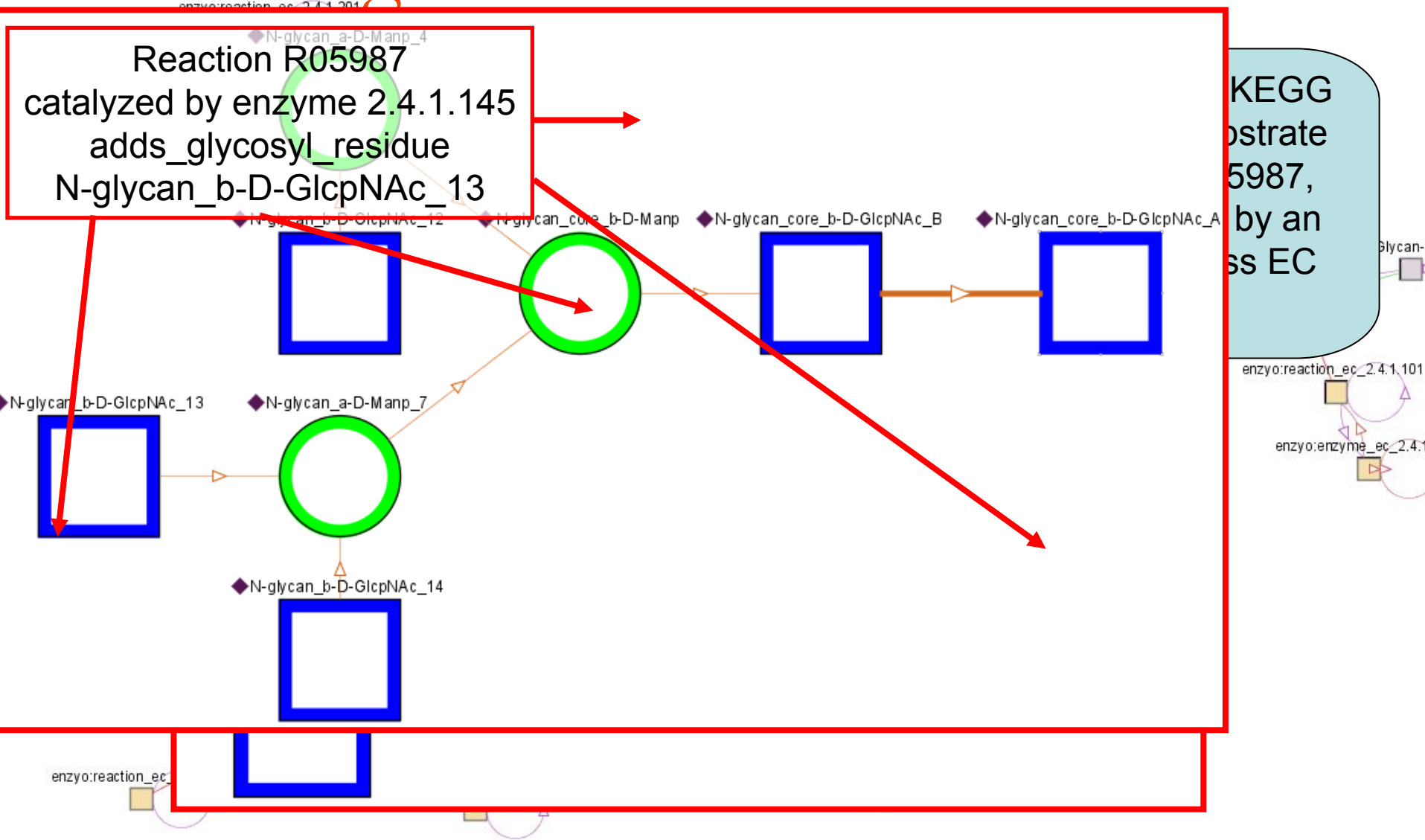


Pathway representation in GlycO

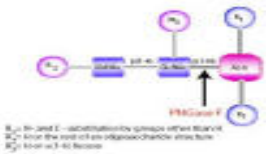


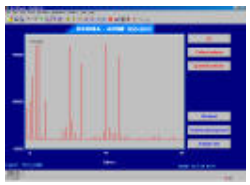
Pathways do not need to be explicitly defined in GlycO. The residue-, glycan-, enzyme- and reaction descriptions contain the knowledge necessary to infer pathways.

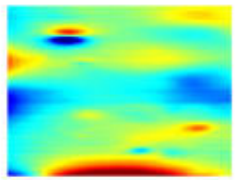
Zooming in a little …



N-Glycosylation Process (NGP)







Cell Culture

extract

Glycoprotein Fraction

proteolysis

Glycopeptides Fraction

Separation technique I

1

n

Glycopeptides Fraction

PNGase

n

Peptide Fraction

Separation technique II

n*m

Peptide Fraction

Mass spectrometry

ms data

Data reduction

ms peaklist

binning

N-dimensional array

Data correlation

Glycopeptide identification and quantification

Signal integration

ms/ms data

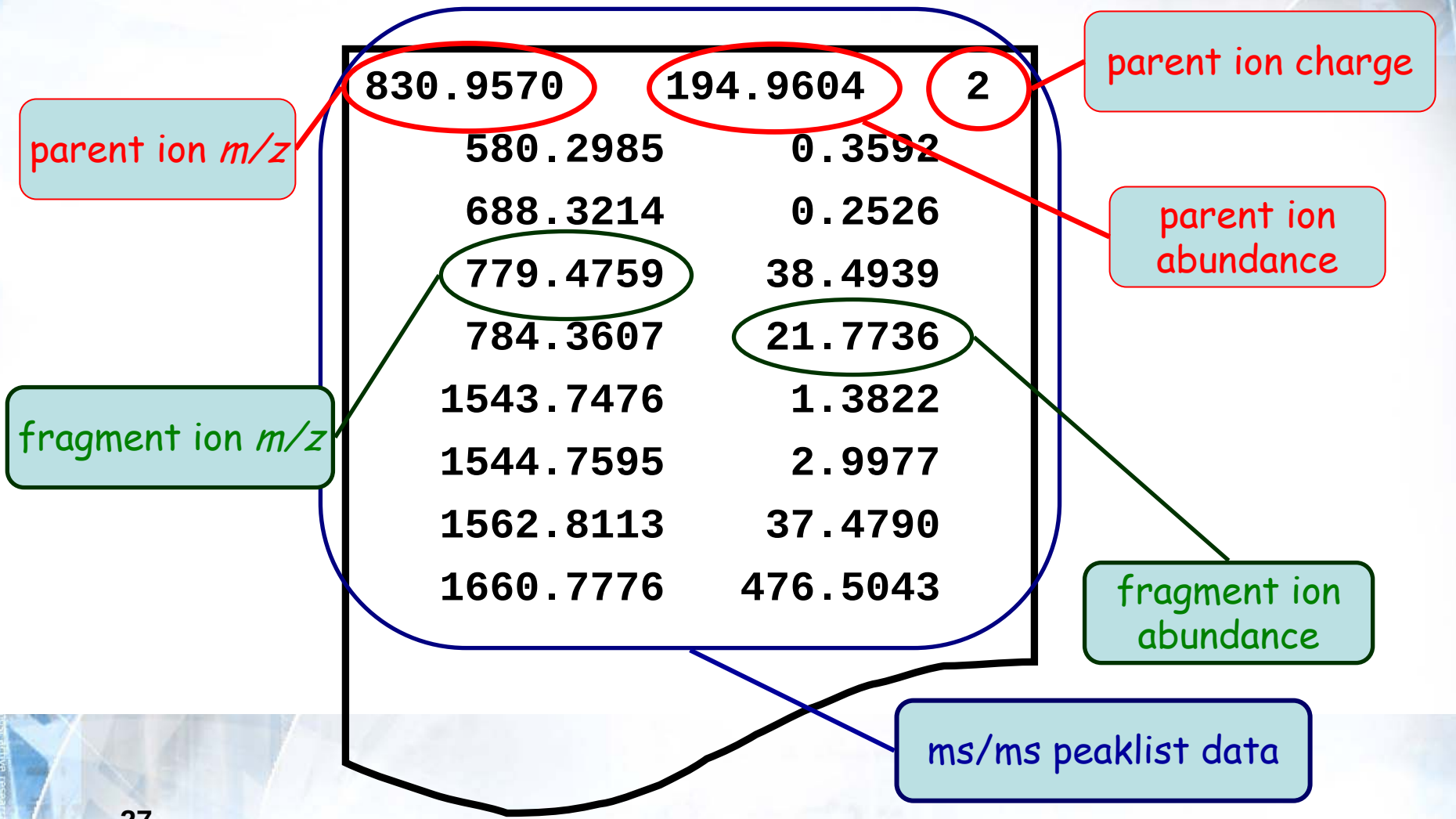
Data reduction

ms/ms peaklist

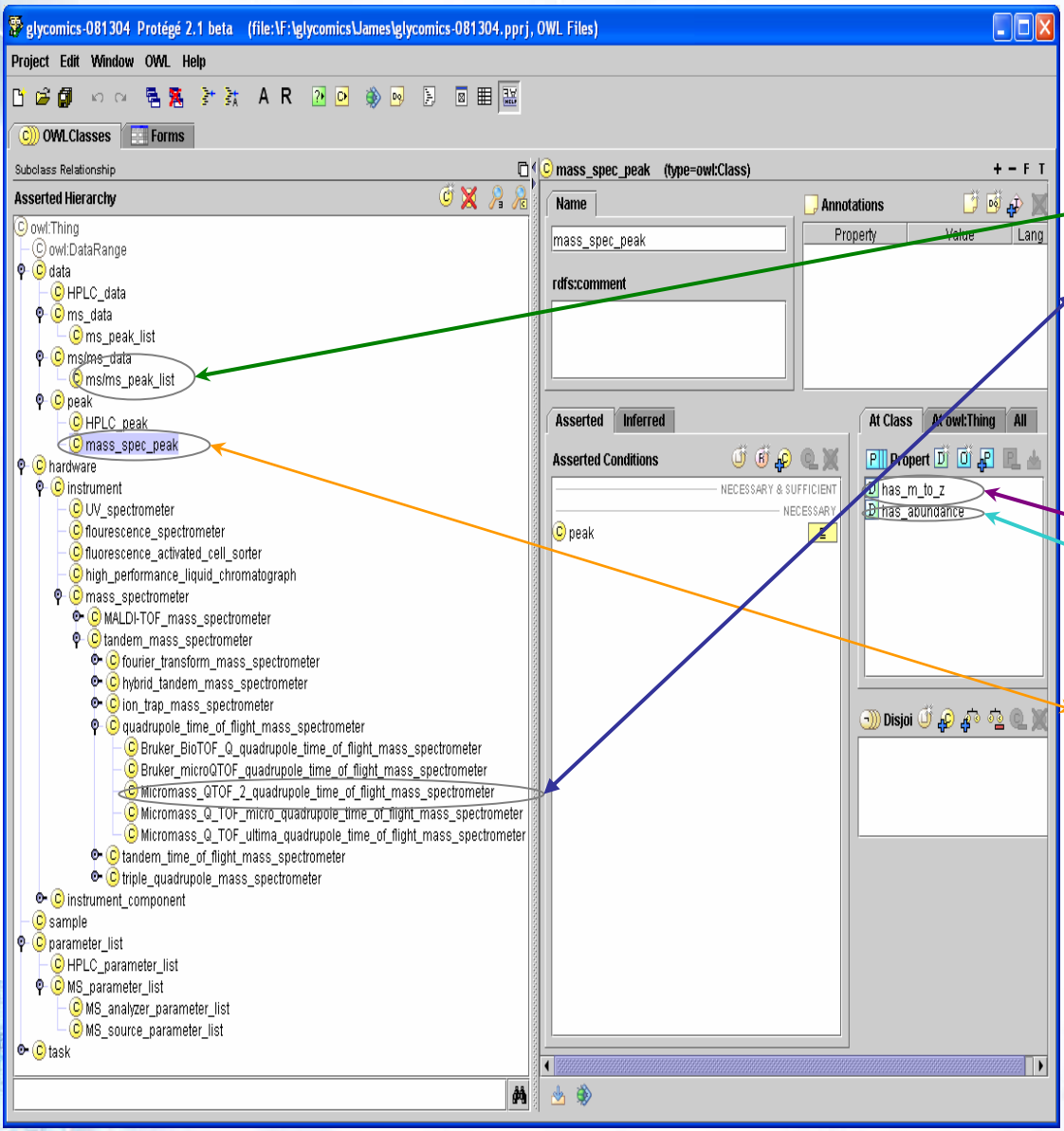
Peptide identification

Peptide list

Semantic Annotation of MS Data



Semantic annotation of Scientific Data



<ms/ms_peak_list>
<parameter
instrument="micromass_QTOF_2_quadropole_time_of_flight_mass_spectrometer"
mode = "ms/ms"/>
<parent_ion_mass>830.9570</parent_ion_mass>
<total_abundance>194.9604</total_abundance>
<z>2</z>
<mass_spec_peak m/z = 580.2985 abundance = 0.3592/>
<mass_spec_peak m/z = 688.3214 abundance = 0.2526/>
<mass_spec_peak m/z = 779.4759 abundance = 38.4939/>
<mass_spec_peak m/z = 784.3607 abundance = 21.7736/>
<mass_spec_peak m/z = 1543.7476 abundance = 1.3822/>
<mass_spec_peak m/z = 1544.7595 abundance = 2.9977/>
<mass_spec_peak m/z = 1562.8113 abundance = 37.4790/>
<mass_spec_peak m/z = 1660.7776 abundance = 476.5043/>
<ms/ms_peak_list>

Annotated ms/ms peaklist data

Summary, Observations, Conclusions

- Deployed health care application that uses SW technologies and W3C recommendations with some understanding of ROI
- New methods for integration and analysis/discovery in biology driven by large populated ontologies
- Projects, library and resources including ontologies at the LSDIS lab: http://lsdis.cs.uga.edu, WWW2006 paper