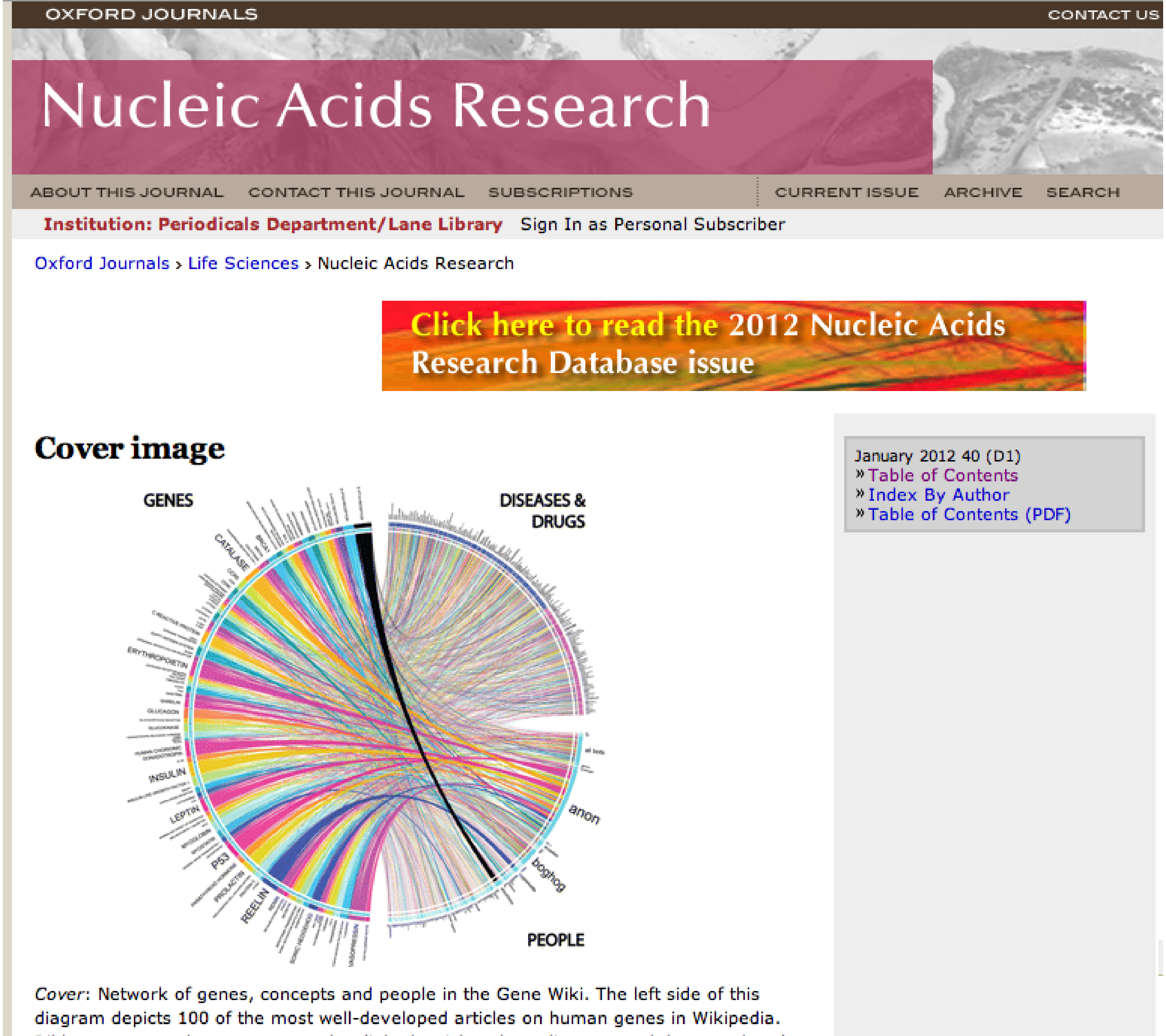
Biomedical Metamodels

Mark A. Musen
Stanford University

NATIONAL CENTER FOR
BIOMEDICAL ONTOLOGY



OXFORD JOURNALS

CONTACT US

Nucleic Acids Research

ABOUT THIS JOURNAL

CONTACT THIS JOURNAL

SUBSCRIPTIONS

CURRENT ISSUE

ARCHIVE

SEARCH

Institution: Periodicals Department/Lane Library Sign In as Personal Subscriber

Oxford Journals › Life Sciences › Nucleic Acids Research

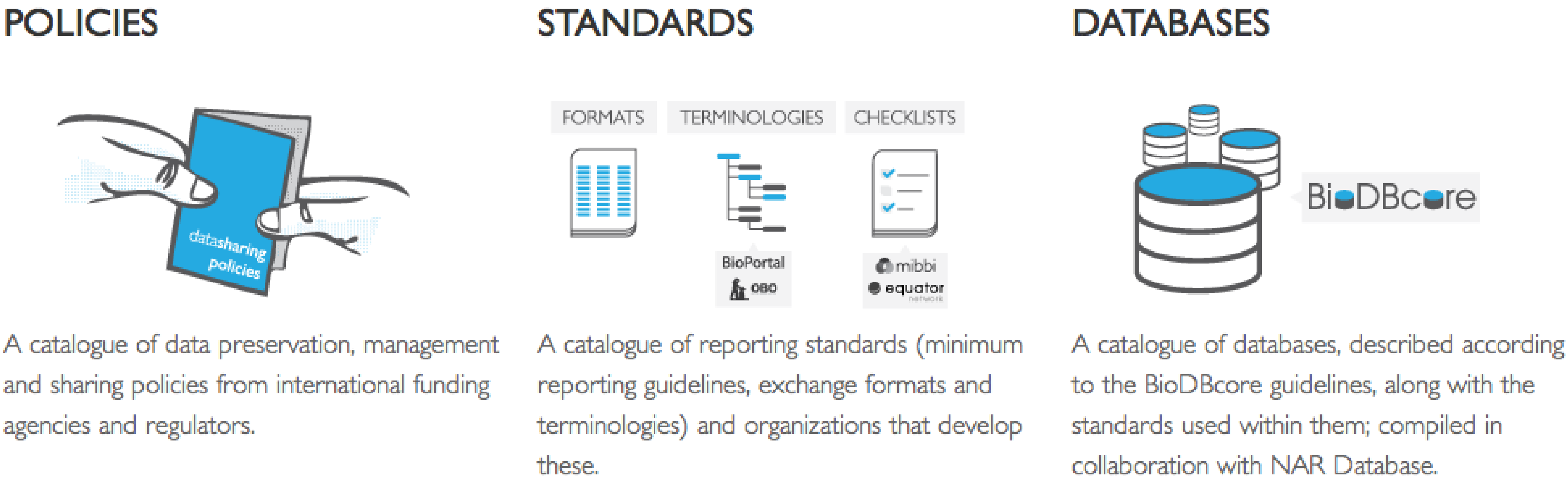
Click here to read the 2012 Nucleic Acids Research Database issue

Cover image

Cover: Network of genes, concepts and people in the Gene Wiki. The left side of this diagram depicts 100 of the most well-developed articles on human genes in Wikipedia.

January 2012 40 (D1)
» Table of Contents
» Index By Author
» Table of Contents (PDF)

BioSharing Initiative attempts to provide a path through all the data—and all the metadata



POLICIES

A catalogue of data preservation, management and sharing policies from international funding agencies and regulators.

STANDARDS

A catalogue of reporting standards (minimum reporting guidelines, exchange formats and terminologies) and organizations that develop these.

DATABASES

A catalogue of databases, described according to the BioDBcore guidelines, along with the standards used within them; compiled in collaboration with NAR Database.

http://biosharing.org

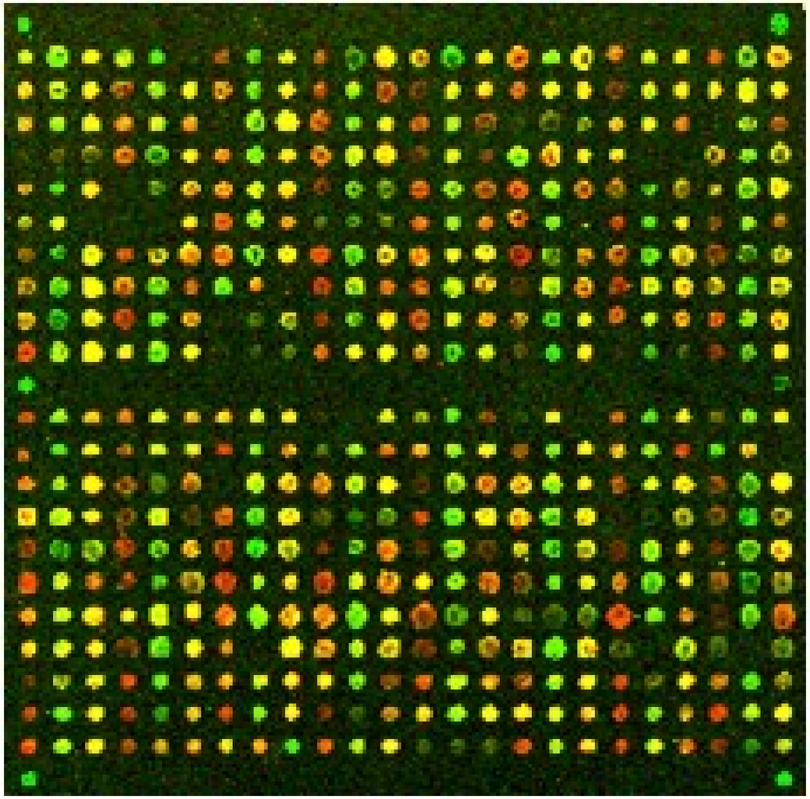
BioDBcore: Uniform descriptors of public biological databases

- Grass-roots effort to develop community-defined descriptions of the core attributes of biomedical databases
- Overseen by the International Society for Biocuration
- Goals are to
  - Overview the landscape of online data resources
  - Encourage consistency and interoperability
  - Promote use of standards

bioDBcore

Microarrays produce tons of gene expression data

- Genes are expressed at different levels in different cells and tissues at different times
- Microarrays bind to specific nucleotide sequences expressed in cells
- Result: Vast amounts of gene-expression data



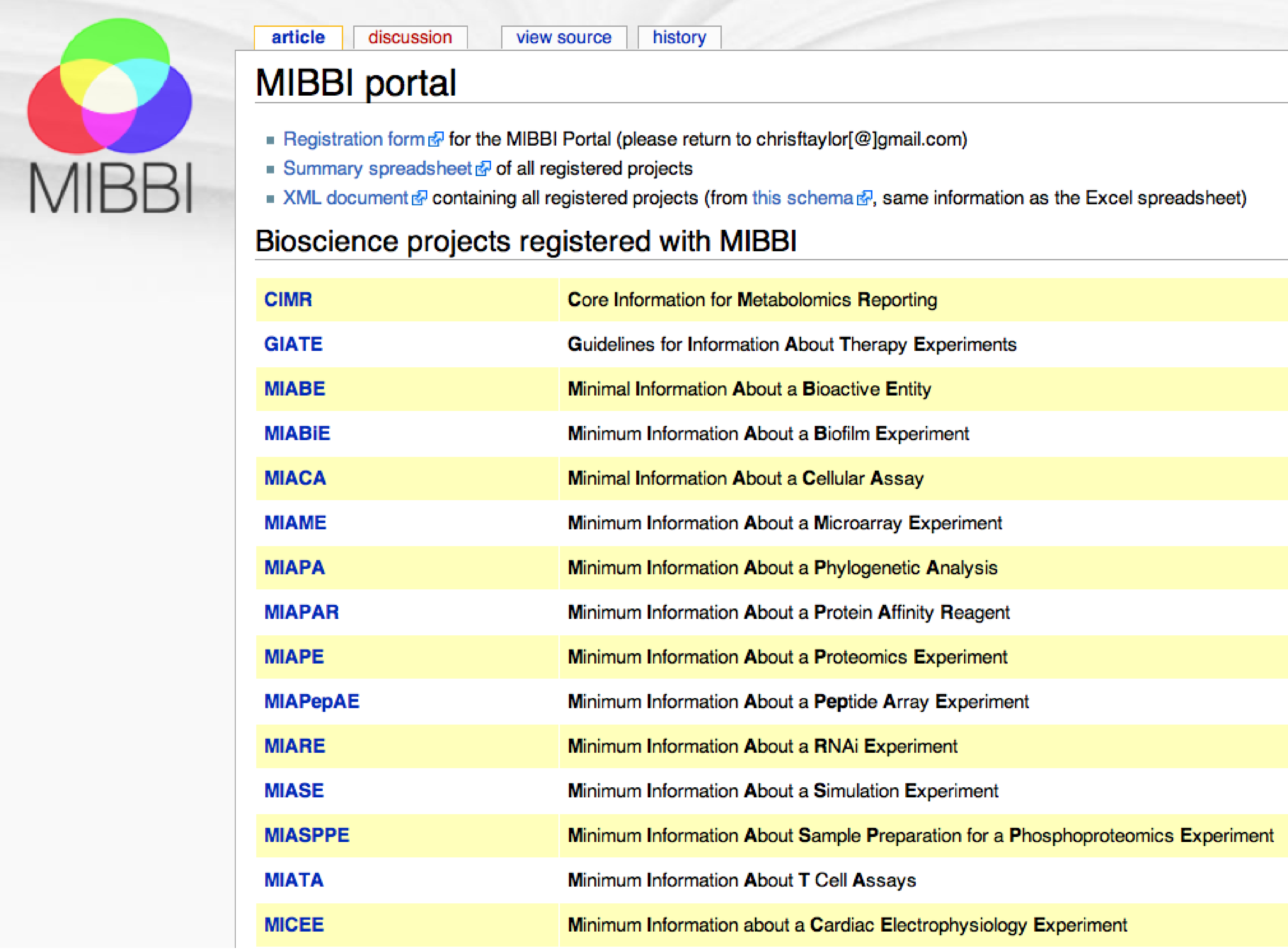
DNA Microarray

Minimum Information About a Microrarray Experiment (MIAME)

- Just a checklist
- At level of detail of what is required in a journal article
- Does not commit to any particular format to enable computational validation or analysis of entries
- Adopted by Gene Expression Omnibus, ArrayExpress and other online databases

Moving beyond checklists

- The Functional Genomics Data Society (formerly the Microarray Gene Expression Data Society; MGED) was founded as the grass-roots effort that created MIAME
- The MicroArray Gene Expression (MAGE) initiative led to
  - **MAGE-ML**: An XML-based markup language
  - **MAGE-TAB**: An Excel-based version of MAGE-ML
  - **MAGE-OM**: A UML-based object language that informs all tags in MAGE-ML
  - **MGED Ontology**: An OWL-based ontology to provide standardized data values of MAGE-OM classes



MIBBI portal

- Registration form for the MIBBI Portal (please return to chrisftaylor[@]gmail.com)
- Summary spreadsheet of all registered projects
- XML document containing all registered projects (from this schema, same information as the Excel spreadsheet)

Bioscience projects registered with MIBBI

CIMR	**C**ore **I**nformation for **M**etabolomics **R**eporting
GIATE	**G**uidelines for **I**nformation **A**bout **T**herapy **E**xperiments
MIABE	**M**inimal **I**nformation **A**bout a **B**ioactive **E**ntity
MIABiE	**M**inimum **I**nformation **A**bout a **Bi**ofilm **E**xperiment
MIACA	**M**inimal **I**nformation **A**bout a **C**ellular **A**ssay
MIAME	**M**inimum **I**nformation **A**bout a **M**icroarray **E**xperiment
MIAPA	**M**inimum **I**nformation **A**bout a **P**hylogenetic **A**nalysis
MIAPAR	**M**inimum **I**nformation **A**bout a **P**rotein **A**ffinity **R**eagent
MIAPE	**M**inimum **I**nformation **A**bout a **P**roteomics **E**xperiment
MIAPepAE	**M**inimum **I**nformation **A**bout a **Pep**tide **A**rray **E**xperiment
MIARE	**M**inimum **I**nformation **A**bout a **R**NAi **E**xperiment
MIASE	**M**inimum **I**nformation **A**bout a **S**imulation **E**xperiment
MIASPPE	**M**inimum **I**nformation **A**bout **S**ample **P**reparation for a **P**hosphoproteomics **E**xperiment
MIATA	**M**inimum **I**nformation **A**bout **T** Cell **A**ssays
MICEE	**M**inimum **I**nformation about a **C**ardiac **E**lectrophysiology **E**xperiment

The HL7 Organization

- ANSI-accredited Standards Development Organization for healthcare applications
- Initial work was applications-layer communication model for data interchange among HIS components
- Current work focuses on
  - Decision support (e.g., Arden Syntax, “Infobuttons”)
  - Enterprise modeling (e.g., Reference Information Model)
  - Clinical Document Architecture, Continuity of Care Documents
  - Clinical terminology

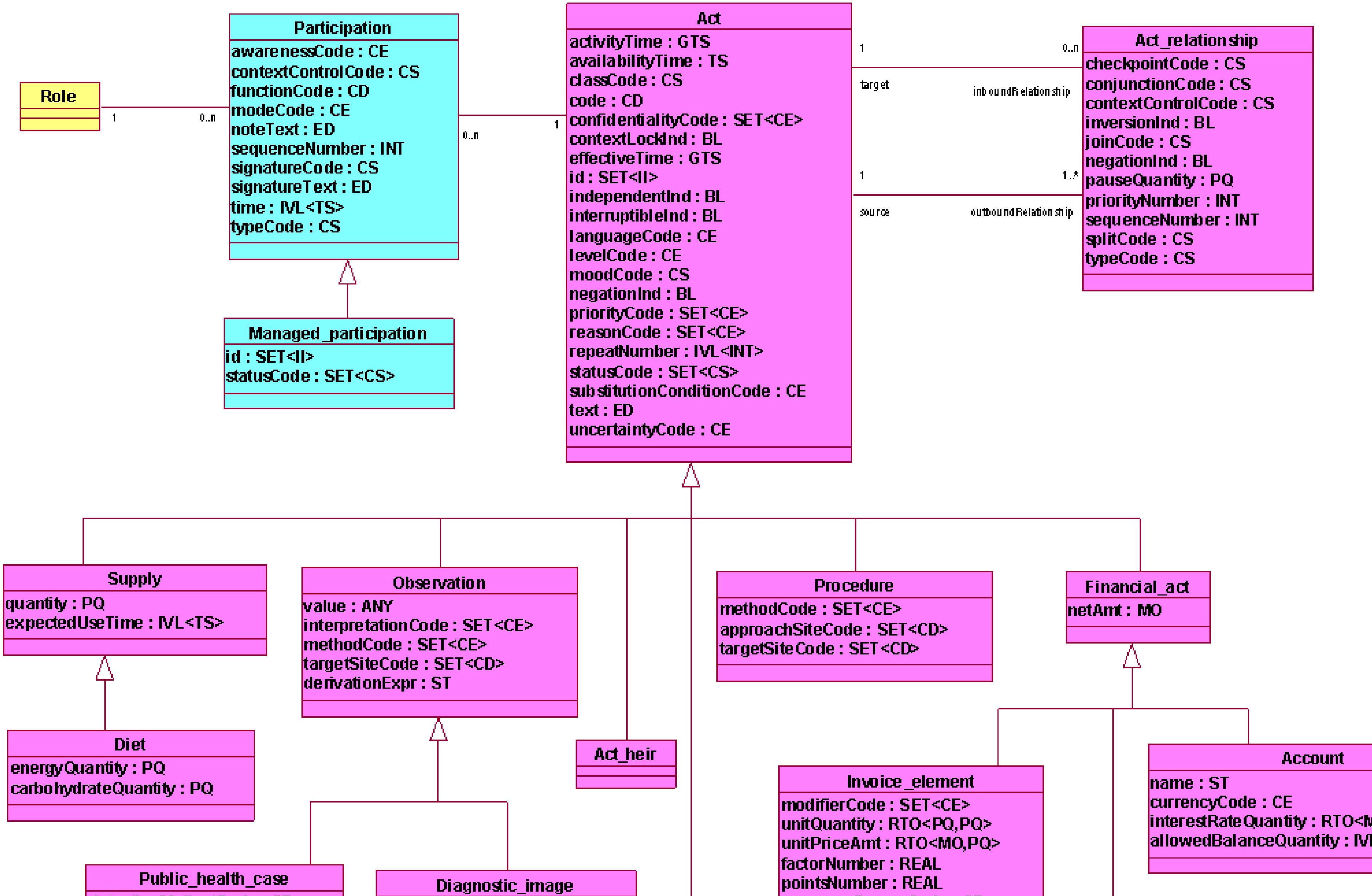
HL7 Version 2 Messaging Standard

```
MSH|^~\&|MegaReg|XYZHospC|SuperOE|XYZImgCtr|20060529090131-
0500||ADT^A01^ADT_A01|01052901|P|2.5  EVN||200605290901||||200605290900  PI
D|||56782445^^^UAReg^PI||KLEINSAMPLE^BARRY^Q^JR||19620910|M||2028-
9^^HL70005^RA99113^^XYZ|260 GOODWIN CREST DRIVE^^BIRMINGHAM^AL^35
209^^M~NICKELL'S PICKLES^10000 W 100TH AVE^BIRMINGHAM^AL^35200^^O
|||||||0105I30001^^^99DEF^AN  PV1||I|W^389^1^UABH^^^^3||||12345^MORGAN^R
EX^J^^^MD^0010^UAMC^L||678
90^GRAINGER^LUCY^X^^^MD^0010^UAMC^L|MED|||||A0||13579^POTTER^SHER
MAN^T^^^MD^0010^UAMC^L|||||||||||||||||||||||||||200605290900  OBX|1|NM|
^Body Height||1.80|m^Meter^ISO+|||||F  OBX|2|NM|^Body
Weight||79|kg^Kilogram^ISO+|||||F  AL1|1||^ASPIRIN  DG1|1||786.50^CHEST PAIN,
UNSPECIFIED^I9|||A
```

HL7 Version 3 Stated Goals

- Apply best practices of software engineering — a model-based methodology
- Adopt three semantic foundations
  1. A reference information model
  2. A robust set of data types
  3. A complete, carefully selected set of terminology domains
- Require all HL7 Version 3 standards to draw from these three common resources

HL-7 Reference Information Model (RIM)



Role

Participation
awarenessCode : CE
contextControlCode : CS
functionCode : CD
modeCode : CE
noteText : ED
sequenceNumber : INT
signatureCode : CS
signatureText : ED
time : IVL<TS>
typeCode : CS

Managed_participation
id : SET<II>
statusCode : SET<CS>

Act
activityTime : GTS
availabilityTime : TS
classCode : CS
code : CD
confidentialityCode : SET<CE>
contextLockInd : BL
effectiveTime : GTS
id : SET<II>
independentInd : BL
interruptibleInd : BL
languageCode : CE
levelCode : CE
moodCode : CS
negationInd : BL
priorityCode : SET<CE>
reasonCode : SET<CE>
repeatNumber : IVL<INT>
statusCode : SET<CS>
substitutionConditionCode : CE
text : ED
uncertaintyCode : CE

Act_relationship
checkpointCode : CS
conjunctionCode : CS
contextControlCode : CS
inversionInd : BL
joinCode : CS
negationInd : BL
pauseQuantity : PQ
priorityNumber : INT
sequenceNumber : INT
splitCode : CS
typeCode : CS

Supply
quantity : PQ
expectedUseTime : IVL<TS>

Diet
energyQuantity : PQ
carbohydrateQuantity : PQ

Observation
value : ANY
interpretationCode : SET<CE>
methodCode : SET<CE>
targetSiteCode : SET<CD>
derivationExpr : ST

Public_health_case

Diagnostic_image

Act_heir

Procedure
methodCode : SET<CE>
approachSiteCode : SET<CD>
targetSiteCode : SET<CD>

Financial_act
netAmt : MO

Invoice_element
modifierCode : SET<CE>
unitQuantity : RTO<PQ,PQ>
unitPriceAmt : RTO<MO,PQ>
factorNumber : REAL
pointsNumber : REAL

Account
name : ST
currencyCode : CE
interestRateQuantity : RTO<M
allowedBalanceQuantity : IV

The HL7 Response to v3 Woes

- Instituted the “Fresh Look” task force in January 2011
- New work on Fast Healthcare Interoperability Resources (FHIR)
  - Still based on Version 3 RIM
  - Adopts “80/20” rule in attempt to simplify things
  - Goal is to make life simpler for implementers, rather than for modelers
  - A modular approach, based on XML

Clinical Document Architecture (CDA)

- First HL7 standard based on the RIM
- XML-based language for marking up narrative clinical reports
- Allows document text to be marked up with terms from the RIM and from controlled terminologies
- Facilitates exchange of clinical documents within and between health-care organizations

Continuity of Care Documents

- CCDs use a refinement of the CDA model that incorporates ideas from yet another standard— ASTM’s Continuity of Care Record (CCR)
- CCD is also an HL7 standard
- Many HIEs use CCDs, as does Microsoft HealthVault
- ONCHIT promotes use of CCD model

REPORT TO THE PRESIDENT
REALIZING THE FULL POTENTIAL OF HEALTH INFORMATION TECHNOLOGY TO IMPROVE HEALTHCARE FOR AMERICANS: THE PATH FORWARD

Executive Office of the President

President's Council of Advisors
on Science and Technology

PCAST Major Conclusions

- "Meaningful use" criteria are necessary first steps
- There must be accelerated progress toward "robust exchange of health information"
- There should be a "universal exchange language" to enable data sharing and an IT infrastructure to support data sharing
- All of this is technologically feasible
- ONC needs to move rapidly in this area

Conclusions

- Metadata initiatives in biomedicine reflect the practical problems of the data deluge
- In the life sciences, there are significant initiatives driven by groups of scientists at the grass roots
- In the clinical area, the principal driving forces have been IT vendors (HL7) and government (CMS, ONC)
- Metadata solutions in biomedicine have at best incomplete penetrance
- Many metadata solutions have been met by open hostility