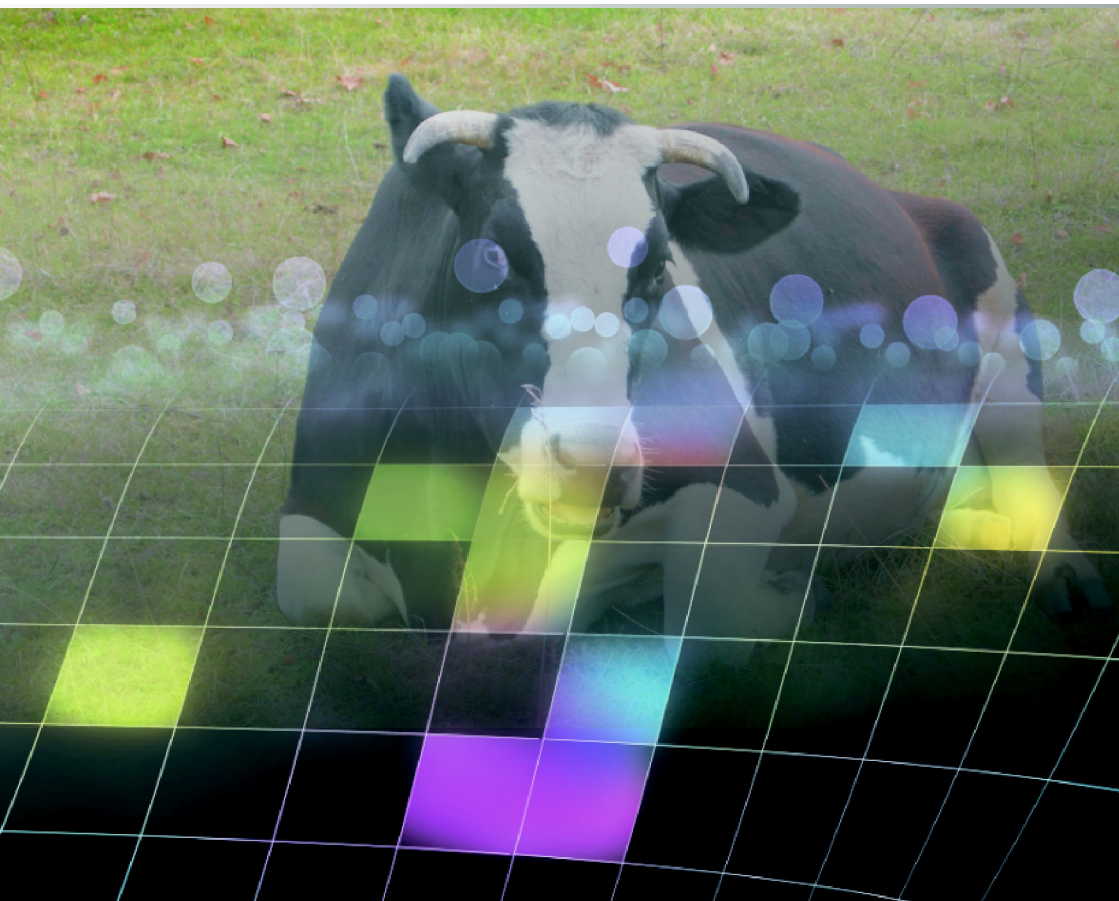


Estrutura básica do MicroArray Gene Expression Markup Language - MAGE-ML



*Empresa Brasileira de Pesquisa Agropecuária
Embrapa Informática Agropecuária
Ministério da Agricultura, Pecuária e Abastecimento*

Documentos 107

Estrutura básica do MicroArray Gene Expression Markup Language - MAGE-ML

Edgard Henrique dos Santos

Embrapa Informática Agropecuária
Campinas, SP
2010

Embrapa Informática Agropecuária

Av. André Tosello, 209 - Barão Geraldo
Caixa Postal 6041 - 13083-886 - Campinas, SP
Fone: (19) 3211-5700 - Fax: (19) 3211-5754
www.cnptia.embrapa.br
sac@cnptia.embrapa.br

Comitê de Publicações

Presidente: *Silvia Maria Fonseca Silveira Massruhá*

Membros: *Poliana Fernanda Giachetto, Roberto Hiroshi Higa,
Stanley Robson de Medeiros Oliveira, Maria Goretti Gurgel Praxedes,
Adriana Farah Gonzalez, Neide Makiko Furukawa*

Membros suplentes: *Alexandre de Castro, Fernando Attique Máximo,
Paula Regina Kuser Falcão*

Supervisor editorial: *Neide Makiko Furukawa*

Revisor de texto: *Adriana Farah Gonzalez*

Normalização bibliográfica: *Maria Goretti Gurgel Praxedes*

Editoração eletrônica/Arte capa: *Suzilei Almeida Carneiro*

Fotos da capa: *Imagens livres disponíveis em <<http://www.stock.schng>>*

Secretária: *Carla Cristiane Osawa*

1ª edição on-line 2010

Todos os direitos reservados.

A reprodução não autorizada desta publicação, no todo ou em parte,
constitui violação dos direitos autorais (Lei no 9.610).

Dados Internacionais de Catalogação na Publicação (CIP) Embrapa Informática Agropecuária

Santos, Edgard Henrique dos.

Estrutura básica do microarray gene expression markup language -
MAGE-ML / Edgard Henrique dos Santos. - Campinas : Embrapa Informática
Agropecuária, 2010.

27 p. : il. – (Documentos / Embrapa Informática Agropecuária ;
ISSN 1677-9274, 107).

1. Microarranjo . 2. MAGE-ML. I. Título. II. Série.

CDD (21. ed.) 572.865

© Embrapa 2010

Autor

Edgard Henrique dos Santos

Analista da Embrapa Informática Agropecuária
Av. André Tosello, 209, Barão Geraldo
Caixa Postal 6041 - 13083-970 - Campinas, SP
Telefone: (19) 3211-5783
e-mail: edgard@cnptia.embrapa.br

Apresentação

Um dos objetivos da Rede Genômica Animal é a identificação de genes que contribuam para o melhoramento de características de interesse econômico em animais de produção.

Uma das ferramentas para prospecção e análise desses genes é o Microarranjo de DNA, uma técnica que permite avaliar a expressão gênica em condições específicas.

Apesar de seu uso amplamente difundido na comunidade científica, os procedimentos e as informações de experimentos nem sempre são padronizados, a despeito dos esforços na criação de uma linguagem padrão como o MAGE-ML.

Este documento visa apresentar o padrão MAGE-ML para aqueles que ainda não se utilizam desse recurso e gostariam de aprender um pouco a respeito.

Kleber Xavier Sampaio de Souza

Chefe Geral

Embrapa Informática Agropecuária

Sumário

Introdução	9
Conjunto de Pacotes e Classes	10
Experiment (Experimento)	10
BioAssay (Ensaio Biológico)	10
ArrayDesign.....	13
DesignElement	13
BioMaterial	16
BioAssayData	16
QuantitationType	16
Array.....	19
BioEvent	19
Protocol	19
AuditAndSecurity.....	19
Description	24
BioAssayDataCluster (HigherLevelAnalysis)	24
Referências	24
Literatura recomendada.....	24

Estrutura básica do MicroArray Gene Expression Markup Language - MAGE-ML

Edgard Henrique dos Santos

Introdução

Várias funções de um laboratório podem ser integradas em um único *chip* de alguns milímetros quadrados. É o que o termo *lab-on-a-chip* (LOC), significa. Um desses “laboratórios” é o *chip* de microarranjos de DNA, que consiste em um conjunto de milhares de moléculas de oligonucleotídeos ligadas de forma sequencial a uma superfície sólida.

Esse *chip* serve como base para a medição dos níveis de expressão de transcritos; análise muito utilizada para a diferenciação, por exemplo, de genes envolvidos em resistência a carrapatos em bovinos ou produção de lã em ovinos.

O MAGE-ML ou *MicroArray Gene Markup Language Expression* é uma linguagem projetada para descrever o conjunto de dados de Microarranjos em um padrão universal.

Baseado em XML, unifica informações de projetos de microarranjos, dados e configuração de experimentos, dados de expressão gênica e análise de resultados.

Este documento descreve a estrutura básica do MAGE-ML usando *Unified Modelling Language* (UML), uma linguagem padrão para descrever modelos de objetos.

O UML foi escolhido pela facilidade de visualização em relação ao XML. Serão mostradas as classes e os pacotes básicos, capazes de descrever um experimento de microarranjos. O pacote completo pode ser acessado pelo site¹.

Conjunto de Pacotes e Classes

Experiment (Experimento)

O experimento é a coleção de todos os ensaios biológicos relacionados ao projeto Experimental (ExperimentDesign).

Esse pacote (Figura 1) descreve um experimento de microarranjo como uma unidade. Ligados a ele temos ensaios biológicos, informações do projeto experimental e um ou mais fatores experimentais que são alterados no curso do experimento para analisar se e como os níveis de expressão gênica se alteraram (por exemplo, com a infestação de carrapatos). Um experimento consiste em um ou mais ensaios biológicos ou etapas de experimentos, e cada ensaio biológico pode testar a expressão de genes para um ou mais valores experimentais fixos (por exemplo, número de carrapatos e tempo de infestação).

BioAssay (Ensaio Biológico)

BioAssay (Ensaio Biológico ou bioensaio) é uma classe abstrata (Figura 2) que representa os agrupamentos físicos e computacionais de arranjos e biomateriais.

O bioensaio é um pequeno passo dentro de um experimento de microarranjo. Existem três tipos de bioensaios: o bioensaio físico, que corresponde ao experimento de microarranjo em laboratório (bancada), o bioensaio medido, que corresponde ao processo após a extração da amostra, e o bioensaio derivado, que corresponde aos passos de processamento dos

¹ Disponível em: <<http://www.ebi.ac.uk/microarray/doc/software/schema/MAGE/MAGE.htm>>.

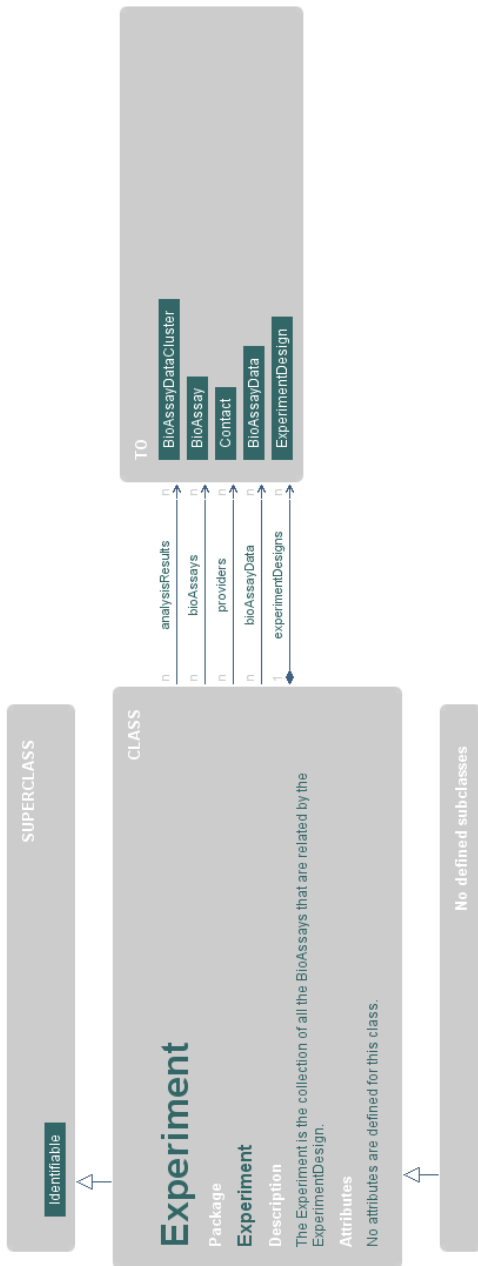


Figura 1. Classe Experiment.

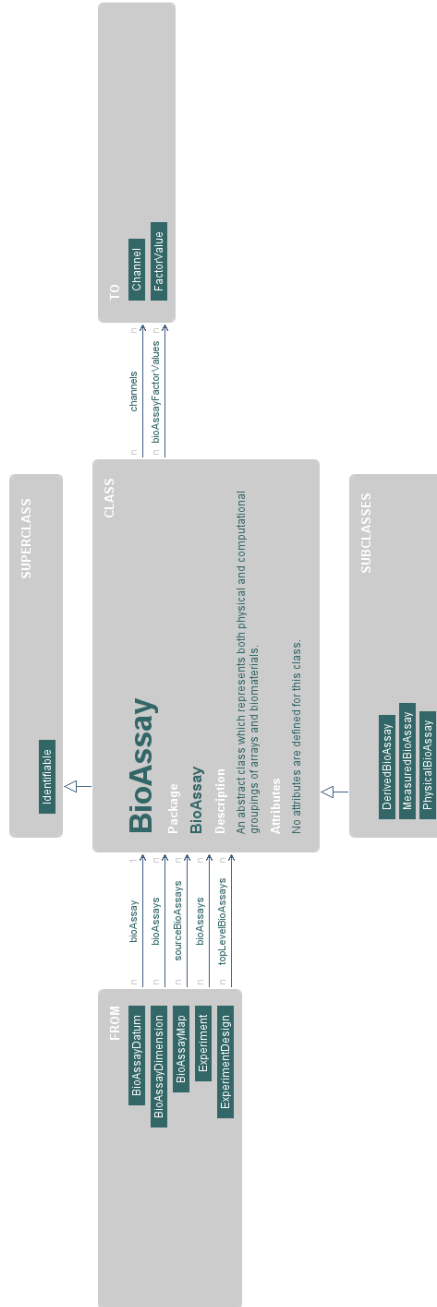


Figura 2. Classe BioAssay.

dados experimentais, ligados à classe `BioAssayMap`.

Um bioensaio físico é criado mediante a aplicação de uma certa quantidade de biomaterial em um *chip* de microarranjo. Os procedimentos durante a execução de um bioensaio (por exemplo, lavar, aplicar agente bloqueador, etc.) podem transformar bioensaios físicos em novos bioensaios físicos. Um tipo específico de procedimento de bioensaio é a aquisição de imagens. Bioensaios medidos podem ser ligados a objetos da classe `MeasuredBioAssayData` (dados brutos).

Os bioensaios derivadas são obtidos pela manipulação de dados e estão ligados pelos objetos da classe `BioAssayMap`.

ArrayDesign

Descreve o layout de uma matriz de expressão gênica e informações sobre elementos de grupos e informações sobre zonas de layout (Figura 3). Existe uma subclasse semelhante, chamada `PhysicalArrayDesign`, que permite a construção de matrizes “virtuais”, as quais descrevem os elementos dos grupos, porém não descrevem as zonas de layout.

As zonas podem formar os chamados “grupos de zonas”. Zonas dentro de um mesmo grupo de zonas teriam o mesmo espaçamento entre elas, enquanto que as zonas de “grupos de zonas” diferentes podem ter diferentes espaçamentos. Zonas e grupos de zonas também podem ser referidos como blocos e metablocos.

DesignElement

Esta classe (Figura 4) representa os elementos de um microarranjo (matriz). São chamados de `probe` ou `CompositeSequence` quando são abstraídas de uma matriz física (*chip*).

O `DesignElement` possui três subclasses. A `CompositeSequence`, que representa o conjunto de probes relacionados à classe `BioSequence` (DNA, RNA ou sequência de proteínas), a classe `Feature`, que contém dados sobre a posição do probe na matriz e a `Probe`, que corresponde à substância fisicamente aderida à matriz.

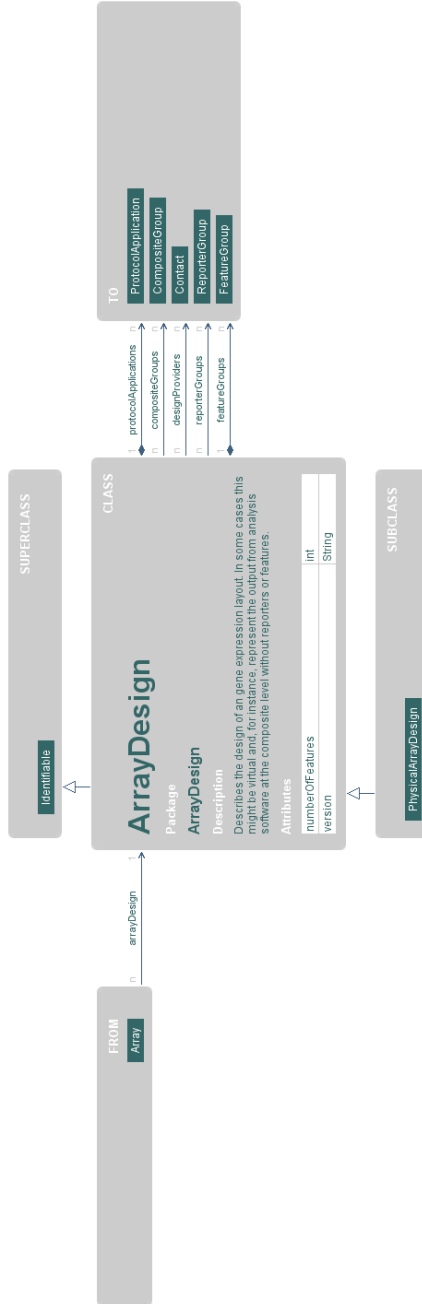


Figura 3. Classe ArrayDesign.

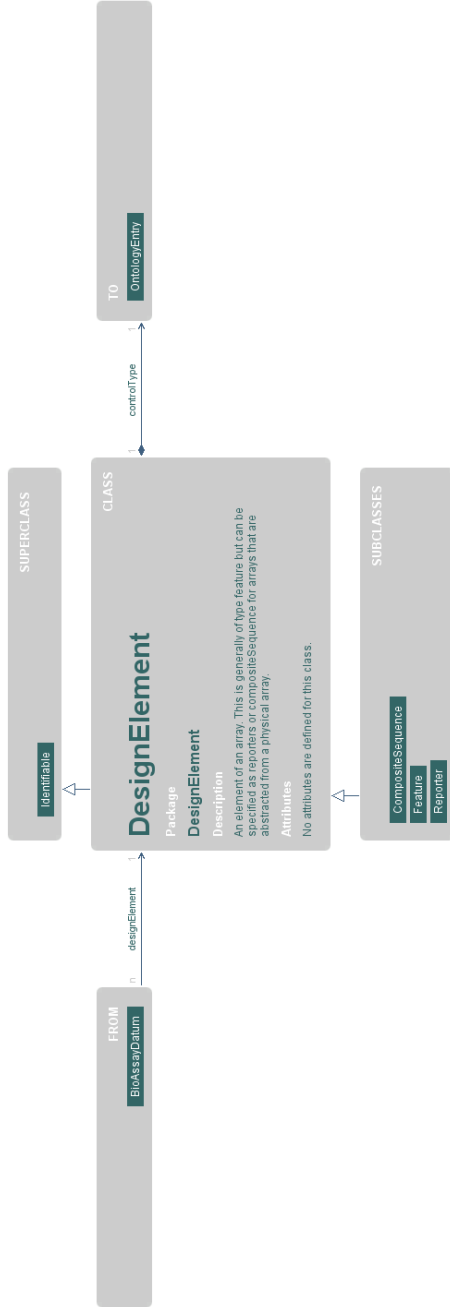


Figura 4. Classe DesignElement.

Para cada probe existe um conjunto de informações que, algumas vezes, podem divergir em relação à probe de referência. Pode haver também uma sequência composta em uma mesma matriz quando há, por exemplo, diferentes variações de splice.

BioMaterial

Os materiais usados nas diversas etapas do experimento de microarranjos são representados nesta classe abstrata chamada BioMaterial (Figura 5). O biomaterial consiste em substâncias como células, tecidos, DNA, proteínas, etc.

A classe apresenta três subclasses. Biosource, que refere-se à fonte inicial do material a ser usado no experimento (por exemplo, célula ou tecido). Biosample refere-se ao material após algum tipo de tratamento. É o último estado do biomaterial antes da hibridação. O biomaterial pode ser resultado de uma cadeia de processos, cada processo com um ou mais biomateriais. Por fim, a classe LabeledExtract, que é um tipo especial da classe BioSamples e que possui, representando as substâncias que são detectáveis, como compostos fluorescentes.

BioAssayData

BioAssayData é a classe que representa o conjunto de dados de bioensaios, sendo o ponto de entrada para estes (Figura 6).

Dados de processos como filtragem ou normalização também podem ser atribuídos a um ou mais objetos dessa classe.

Ela apresenta também duas subclasses: DerivedBioAssayData e MeasuredBioAssayData. A primeira representa dados de saída de um determinado processamento e a segunda, informações sobre o procedimento inicial de um bioensaio.

QuantitationType

QuantitationType descreve os métodos para análise de informações dos dados da matriz de microarranjo. É possível inserir dados de erro ou intensidade de sinais e intensidade de cores (Figura 7).

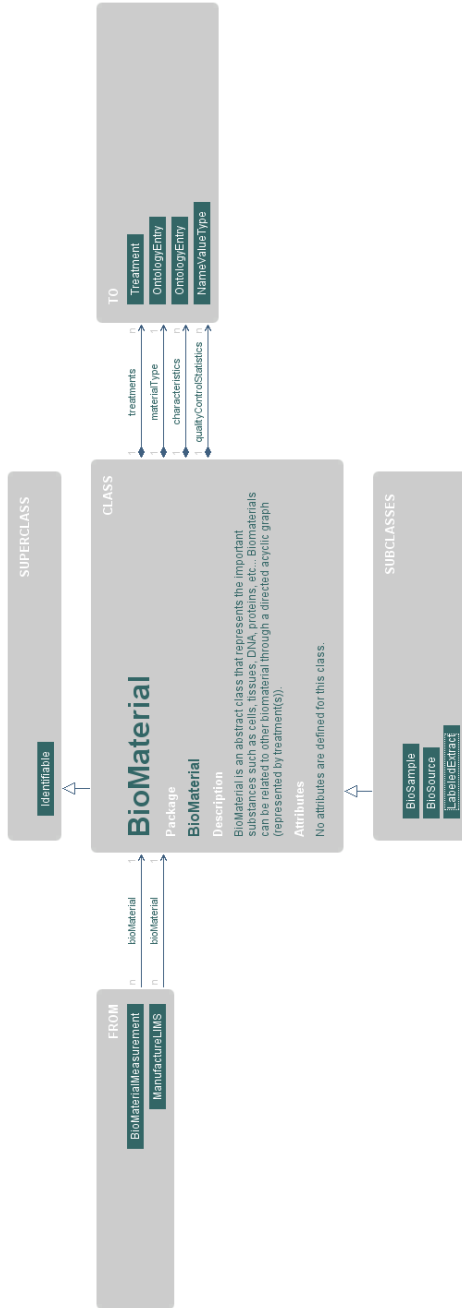


Figura 5. Classe BioMaterial.

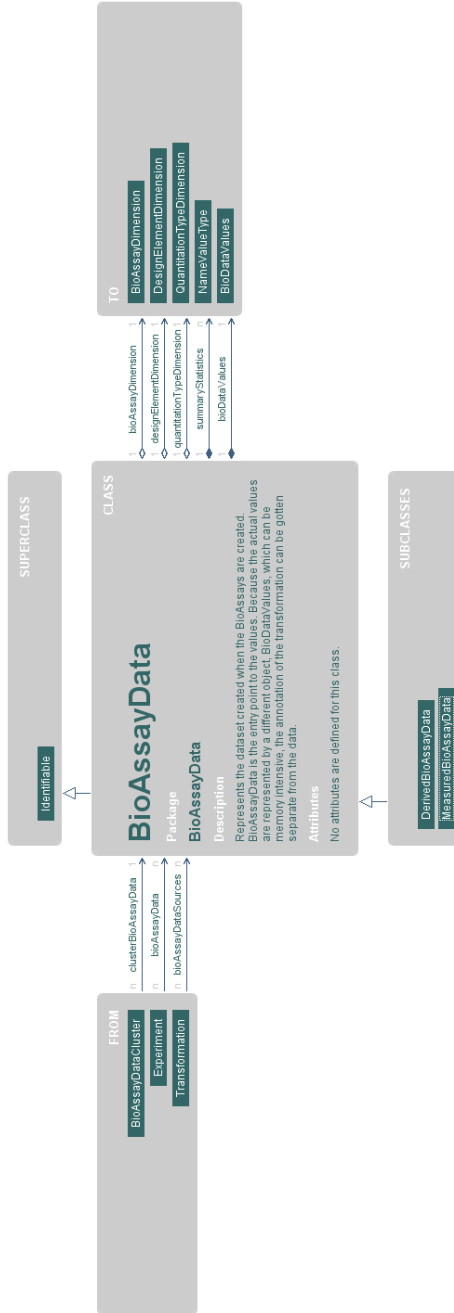


Figura 6. Classe BioAssayData.

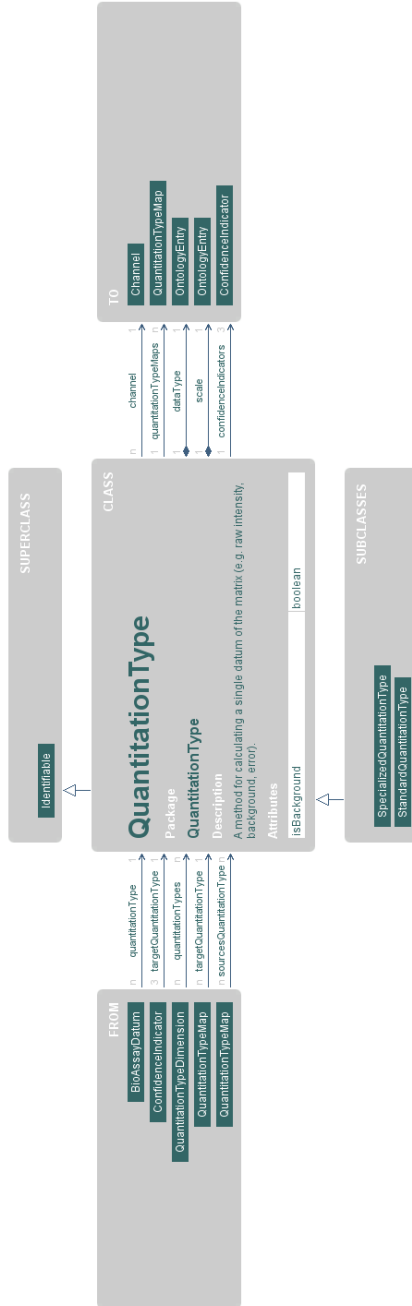


Figura 7. Classe QuantitationType.

Array

Representa a matriz física do microarranjo (Figura 8). Há dois tipos de atributos importantes que podem ser incluídos: A identificação do ou lâmina (fabricante, organismo, etc), as posições X e Y na matriz.

BioEvent

Essa classe (Figura 9) descreve os tipos de eventos possíveis para cada microarranjo. Cada evento pode ser uma sequência de processos ou protocolo.

Existem eventos físicos, como tratamento do biomaterial, criação de um bioensaio, hibridação ou processamentos de bioensaios, e eventos de tratamento de informações, como a obtenção de características, dados ou mapas.

Protocol

Esse pacote (Figura 10) está dividido em duas partes. A primeira é a classe hardware e a segunda software, ambas abstratas.

Um hardware específico pode demandar um software específico, e este pode ser composto por outros softwares ou módulos. As subclasses no segundo nível representam a aplicação dessas entidades abstratas e seus possíveis parâmetros.

AuditAndSecurity

O pacote AuditAndSecurity (Figura 11) permite que as informações de contatos sejam criadas e rastreadas para cada objeto. Um contato pode ser uma organização ou uma pessoa. Uma pessoa pode trabalhar para uma organização, e uma organização pode ser composta por outras suborganizações.

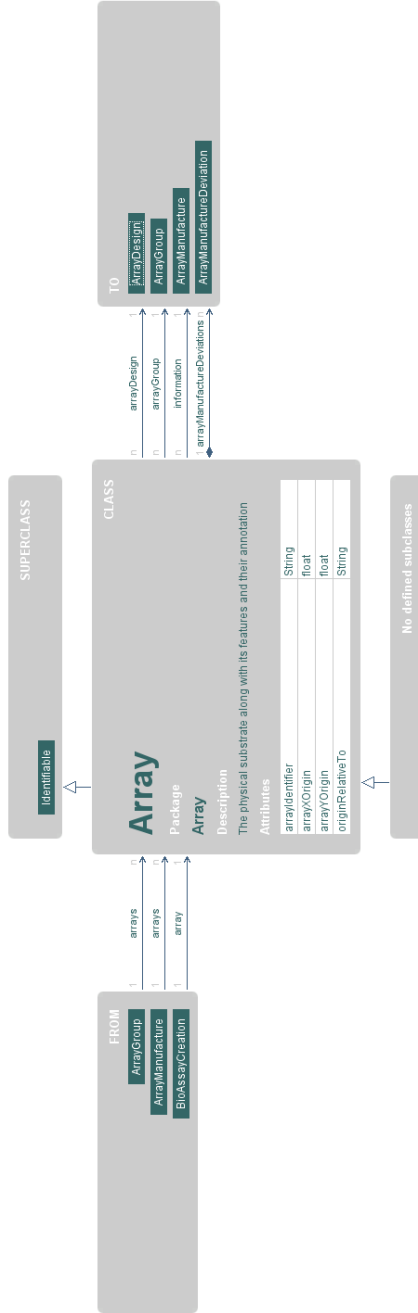


Figura 8. Classe Array.

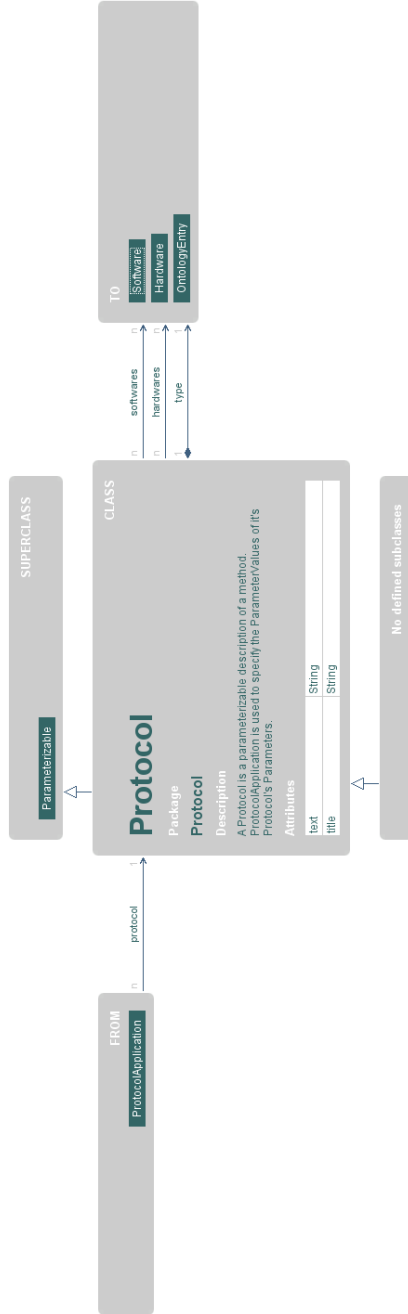


Figura 10. Classe Protocol.

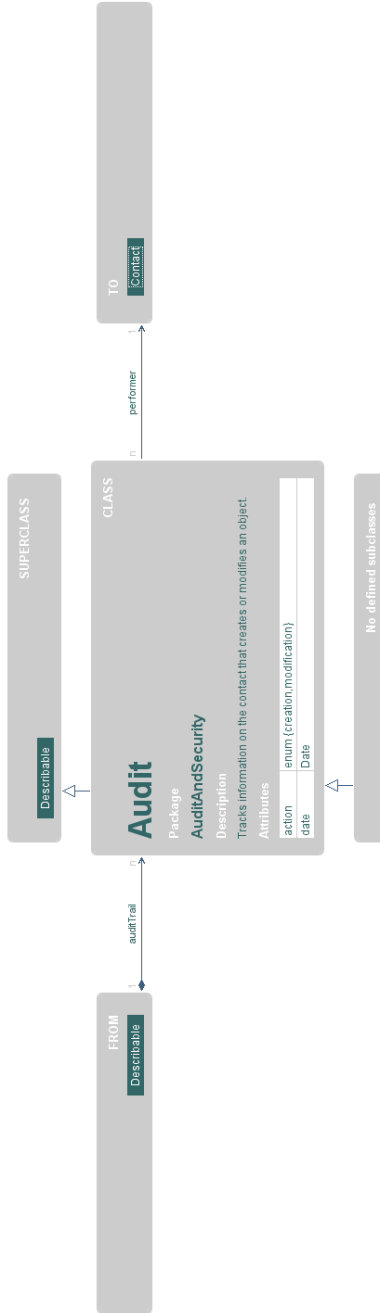


Figura 11. Classe Audit.

Description

Caso não seja possível a inclusão de mais informações aos objetos MAGE, além do escopo dos atributos e relacionamentos, o pacote Description (Figura 12) pode ser utilizado.

Uma descrição pode consistir de um texto livre, referências para as entradas no banco de dados, referências para ontologias (anotações) e uma ou mais referências. Existe ainda classe genérica NameValueType, caso Description ainda não seja suficiente.

BioAssayDataCluster (HigherLevelAnalysis)

BioAssayDataCluster (Figura 13) representa os dados experimentais que podem ser agrupados, obtendo-se um ou mais níveis superiores de *clusters*. Cada *cluster* é composto por nós, e cada nó pode conter subnós. Um nó pode ser caracterizado por seus valores (por exemplo, alguns métricas de qualidade do *cluster*), e grupo de nós podem identificar condições experimentais ou elementos, como *spots* ou genes.

A subclasse BioAssayDimension representa a lista ordenada de bioensaios (condições experimentais) e a DesignElementDimension representa a lista ordenada de elementos. Ambas fazem parte do pacote HigherLevelAnalysis.

Referências

MGED - Microarray gene expression data society. Disponível em: <<http://www.mged.org/Workgroups/MAGE/mage.html>>. Acesso em: 2 nov. 2010.

Literatura recomendada

SPELLMAN, P. T.; MILLER, M.; STEWART, J.; TROUP, C.; SARKANS, U.; CHERVITZ, S.; BERNHART, D.; SHERLOCK, G.; BALL, C.; LEPAGE, M.; SWIATEK, M.; MARKS, W. L.; GONCALVES, J.; MARKEL, S.; IORDAN, D.; SHOJATALAB, M.; PIZARRO, A.; WHITE, J.; HUBLEY, R. DEUTSCH, E.; SENGER, M.; ARONOW, B. J.; ROBINSON, A.; BASSETT, D.; STOECKERT JUNIOR, C. J.; BRAZMA, A. Design and implementation of microarray gene expression markup language (MAGE-ML). *Genome Biology*, v. 3, n. 9, Aug. 2002. research0046.1–0046.9.

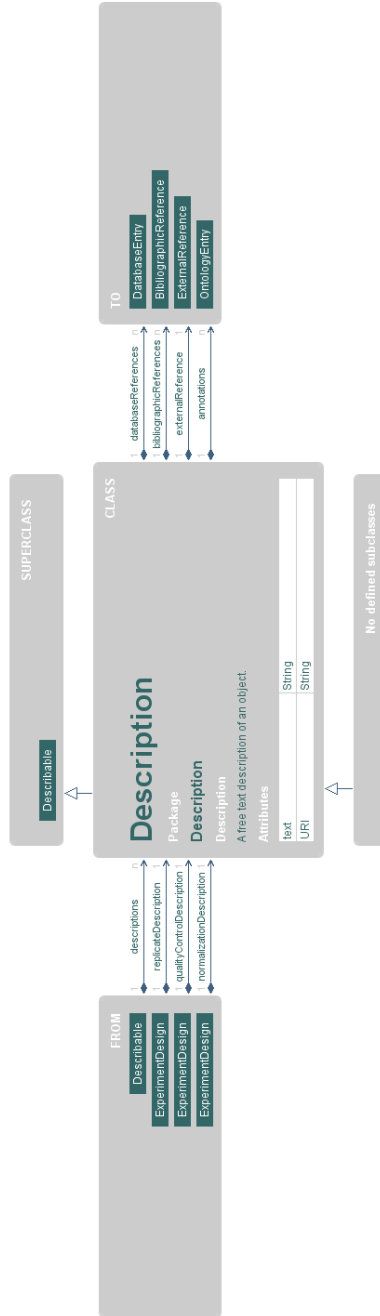


Figura 12. Class Description.



Figura 13. Classe BioAssayDataCluster.

Embrapa

Informática Agropecuária

Ministério da
Agricultura, Pecuária
e Abastecimento



CGPE 9090