ArMet design

Design Status: Draft Release
Design Version: Version 3.1
Date: 7th of June, 2004
Author: Helen Fuell

Department of Computer Science
University of Wales
Aberystwyth
Ceredigion
SY23 3DB
UK

Table of Contents

1. Introduction

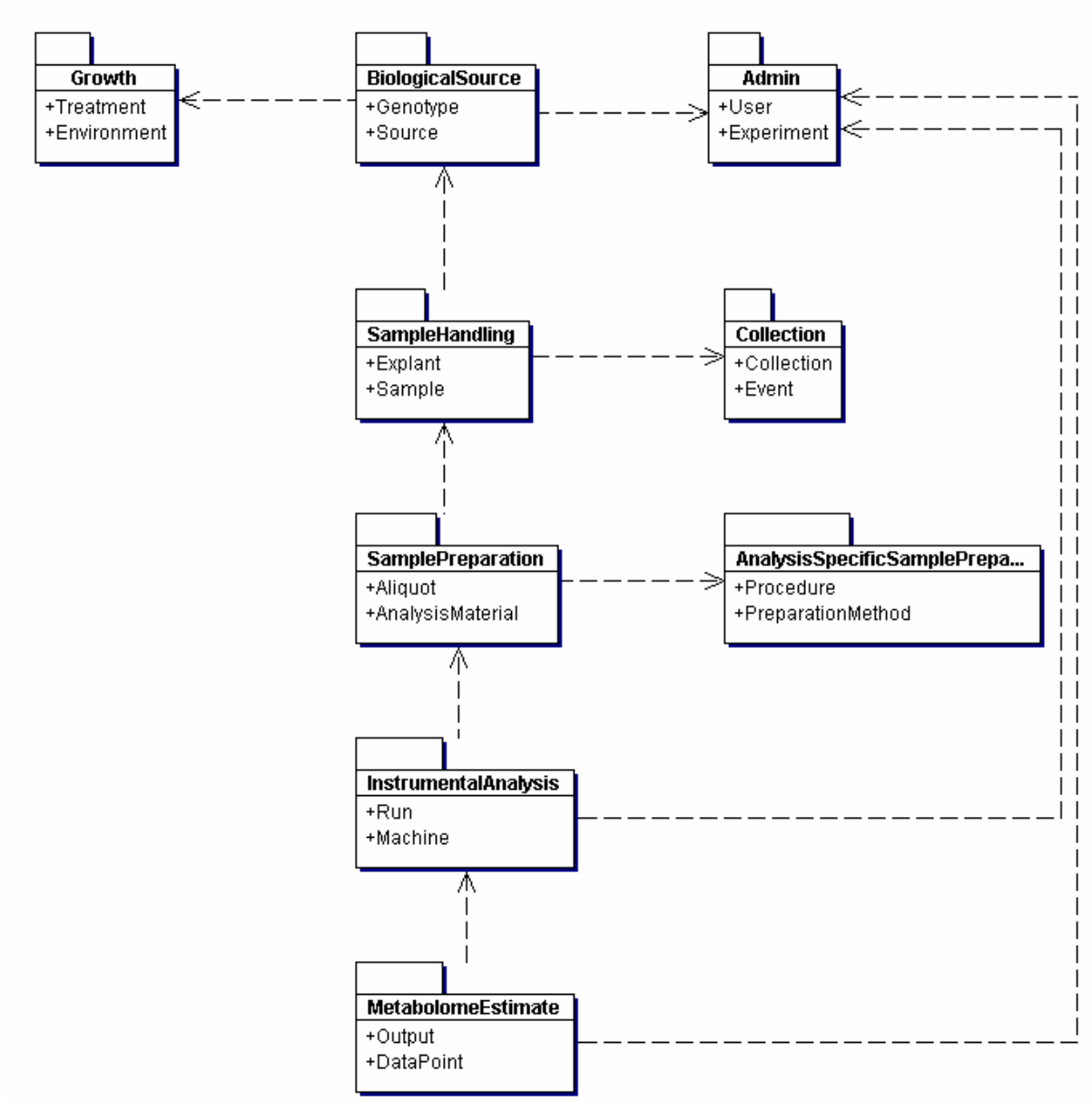
This document presents the design of ArMet through a description of the overall architecture and a presentation of the logical data model for its core.

2. The Architecture

ArMet describes the timeline for a metabolomics experiment by way of nine components which are represented in Figure 1 by way of UML packages and which are described briefly in the list below:

1. **Admin.** Data that identify experiments and the users involved in them.
2. **Biological Source.** Genotype, provenance and identification data that describe the biological source material used in metabolomics experiments.
3. **Growth.** Data that describe the environments in which the biological source material is developed.
4. **Collection.** Data that describe the gathering of samples from biological source material.
5. **Sample Handling.** Data that describe the bulking/division and storage of material gathered during a collection event.
6. **Sample Preparation.** Data that describe the preparation of samples for presentation to an analytical instrument.
7. **Analysis Specific Sample Preparation.** Data that describe analysis/instrument specific sample preparation procedures.
8. **Instrumental Analysis.** Data that describe the analysis of the metabolite content of samples by analytical instruments.
9. **Metabolome Estimate.** Results sets for samples together with associated data pre-processing protocol information.

Figure 1. The ArMet Architecture



The arrows between the packages represent the existence of dependencies between the classes that they contain. When the interface to a class is changed then any classes which depend on that interface may also have to change. The dependency arrows, therefore, depict the relationships between the components and show how they are linked to represent the entire experimental timeline. More detail on the dependencies shown in Figure 1 is provided with the detailed descriptions of each component in Section 4.

Each of the packages in Figure 1 contains classes which represent the key entities about which data is stored for the components. These classes describe a core set of public data items for each entity which should be available for all plant metabolomics experiments. Sub-packages may then be created for each component that contain classes that extend the core classes (and optionally additional classes) to describe a particular experimental methodology or to provide additional information for a particular experiment or project. The core data for each component is described with the detailed descriptions of each component in Section 3.

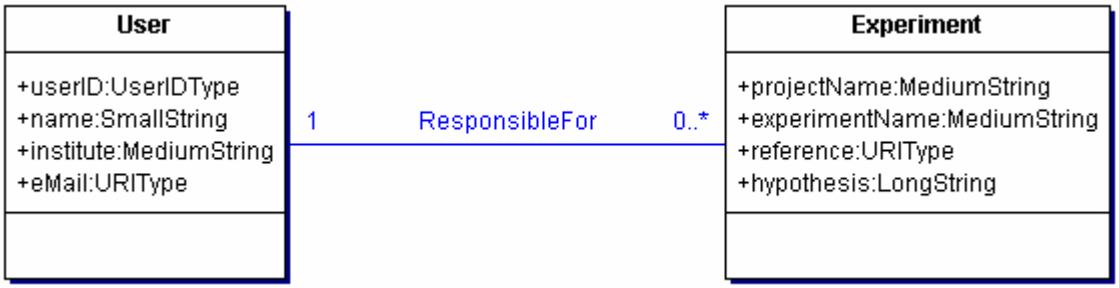
4. The Components

This section will present the design for each component in turn.

4.1. Admin

Class Diagram. The class diagram for the Admin package is depicted in Figure 2.

Figure 2. The Admin Component



Entities. The Admin component contains the following entities:

Table 1. Admin Component Entities

Name	Description
User	An ArMet user; someone involved in a metabolomics experiment.
Experiment	A metabolomics experiment.

Dependencies. The Admin component has no dependencies on other components.

Attributes. The core data items for the Admin component are described below. All of these data items are required. Their data types (as given in Figure 2) are described in Appendix A.

Table 2. Admin Component User Entity Attributes

Attribute	Description
userID	The userID of a valid user of an ArMet implementation. (Primary Key)
name	The given name of the user.
institute	The name of the institute to which the user is affiliated.
eMail	The email address of the user.

Table 3. Admin Component Experiment Entity Attributes

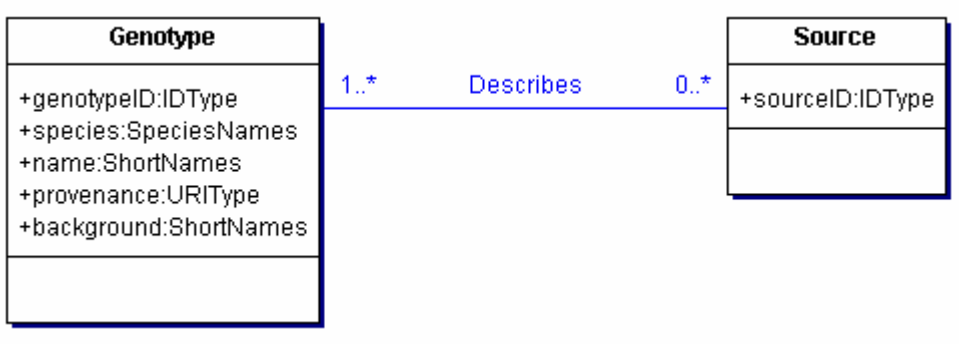
Attribute	Description
projectName	The identifier of the project for which the experiment is carried out. (Partial Primary Key)
experimentName	The identifier of the experiment. (Partial Primary Key)

Attribute	Description
reference	A reference to further information on the experiment.
hypothesis	The question that the experiment aims to answer.

4.2. Biological Source

Class Diagram. The class diagram for the Biological Source package is depicted in Figure 3.

Figure 3. The Biological Source Component



Entities. The Biological Source component contains the following entities:

Table 4. Biological Source Component Entities

Name	Description
Genotype	An ecotype, mutant or transgenic.
Source	An item of biological source material used in an experiment.

Dependencies. The Biological Source component is dependent upon the Admin component and the Growth component. These dependencies can be described by way of the following relationships between entities:

Table 5. Biological Source Component Dependencies

Entities	Multiplicity	Relationship
Experiment:Source	1..*:0..*	Involves
Treatment:Source	1..1:0..*	Involves

Attributes. The core data items for the Biological Source component are described below. All of these data items are required. Their data types (as given in Figure 3) are described in Appendix A.

Table 6. Biological Source Component Genotype Entity Attributes

Attribute	Description
genotypeID	A unique identifier for the genotype. (Primary Key)
species	The latin name of the genotype.

Attribute	Description
name	Short name for the ecotype, mutant or transgenic.
provenance	The source of the genotype.
background	The background of mutants and transgenics.

Additional requirements and constraints. The following additional requirements and constraints exist for the Genotype entity attributes:

- The five letter entity identifier that makes up part of genotypeID should be "genot".
- The values for name and background should be appropriate to the specified species.

Table 7. Biological Source Component Source Entity Attributes

Attribute	Description
sourceID	A unique identifier for the item of source material. (Primary Key)

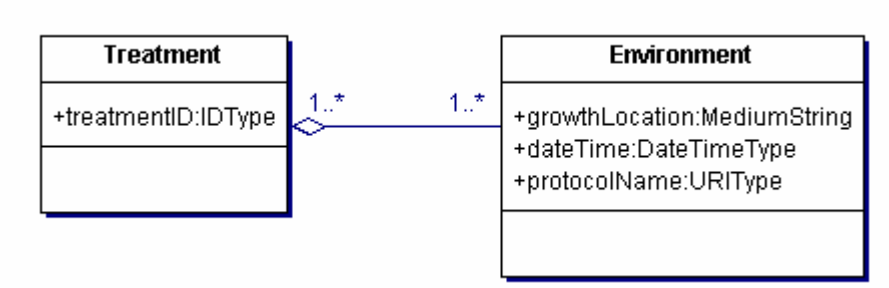
Additional requirements and constraints. The following additional requirements and constraints exist for the Source entity attributes:

- The five letter entity identifier that makes up part of sourceID should be "sourc".

4.3. Growth

Class Diagram. The class diagram for the Growth package is depicted in Figure 4.

Figure 4. The Growth Component



Entities. The Growth component contains the following entities:

Table 8. Growth Component Entities

Name	Description
Treatment	The cultivation of a set of items of biological source material.
Environment	A growth environment applied to a set of items of biological source material undergoing a specified treatment.

Dependencies. The Growth component has no dependencies on other components.

Attributes. The core data items for the Growth component are described below. All of these data items are required. Their data types (as given in Figure 4) are described in Appendix A.

Table 9. Growth Component Treatment Entity Attributes

Attribute	Description
treatmentID	A unique identifier for the treatment. (Primary Key)

Additional requirements and constraints. The following additional requirements and constraints exist for the Treatment entity attributes:

- The five letter entity identifier that makes up part of treatmentID should be "treat".

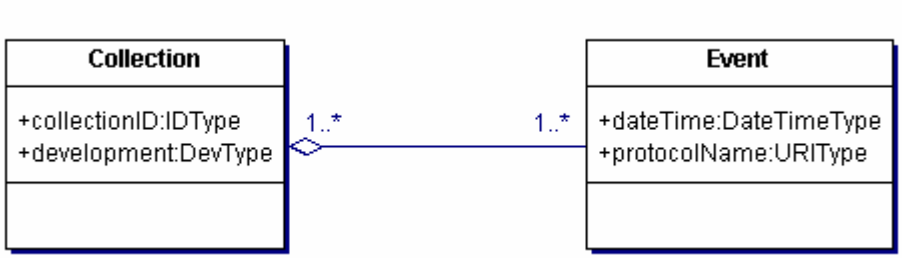
Table 10. Growth Component Environment Entity Attributes

Attribute	Description
growthLocation	The location in which a set of items of biological source material undergo part or all of a treatment. (Partial Primary Key)
dateTime	The date and time at which the environmental conditions were applied. (Partial Primary Key)
protocolName	A reference to the protocol followed in this environment. (Partial Primary Key)

4.4. Collection

Class Diagram. The class diagram for the Collection package is depicted in Figure 5.

Figure 5. The Collection Component



Entities. The Collection component contains the following entities:

Table 11. Collection Component Entities

Name	Description
Collection	The sampling of a set of items of biological source material that are at the same stage of development.
Event	A sampling activity applied to a set of items of biological source material undergoing a specified collection.

Dependencies. The Collection component has no dependencies on other components.

Attributes. The core data items for the Collection component are described below. All of these items are required. Their data types (as given in Figure 5) are described in Appendix A.

Table 12. Collection Component Collection Entity Attributes

Attribute	Description
collectionID	A unique identifier for the collection (Primary Key)
development	The development stage of the items of source material at the time of sampling.

Additional requirements and constraints. The following additional requirements and constraints exist for the Collection entity attributes:

- The five letter entity identifier that makes up part of collectionID should be "clect".
- The DevScale of the development attribute for a particular collection should be appropriate to the genotype of the items of biological source material involved in the collection.
- The date order for multiple collections from the same item of biological source material should be the same as the DevStage order of the development attributes for those collections. The date order of collections may be derived from the dateTime attributes in the events associated with the collections.

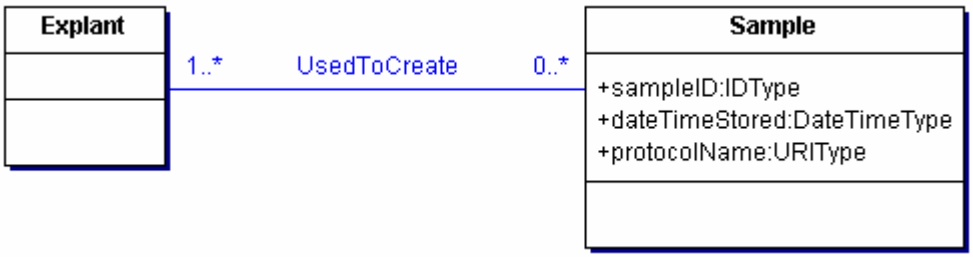
Table 13. Collection Component Event Entity Attributes

Attribute	Description
dateTime	The date and time of the collection activity. (Partial Primary Key)
protocolName	A reference to the protocol followed for this activity. (Partial Primary Key)

4.5. Sample Handling

Class Diagram. The class diagram for the Sample Handling package is depicted in Figure 6.

Figure 6. The Sample Handling Component



Entities. The Sample Handling component contains the following entities:

Table 14. Sample Handling Component Entities

Name	Description
Explant	The material gathered from a single item of biological source material during a collection.
Sample	The result of any bulking and/or division of explants.

Dependencies. The Sample Handling component is dependent upon the Biological Source component and the Collection component. These dependencies can be described by way of the following relationships between entities:

Table 15. Sample Handling Component Dependencies

Entities	Multiplicity	Relationship
Source:Explant	1..1:0..*	Becomes
Collection:Explant	1..1:0..*	Produces

Attributes. The core data items for the Sample Handling component are described below. All of these data items are required. Their data types (as given in Figure 6) are described in Appendix A.

Sample Handling Component Explant Entity Attributes. The Explant entity does not contain any core data items. Its composite primary key comprises the identifiers for the associated source and collection entities.

Table 16. Sample Handling Component Sample Entity Attributes

Attribute	Description
sampleID	A unique identifier for the sample. (Primary Key)
dateTimeStored	The date and time at which the sample was put into storage.
protocolName	A reference to the creation and storage protocol followed for this sample.

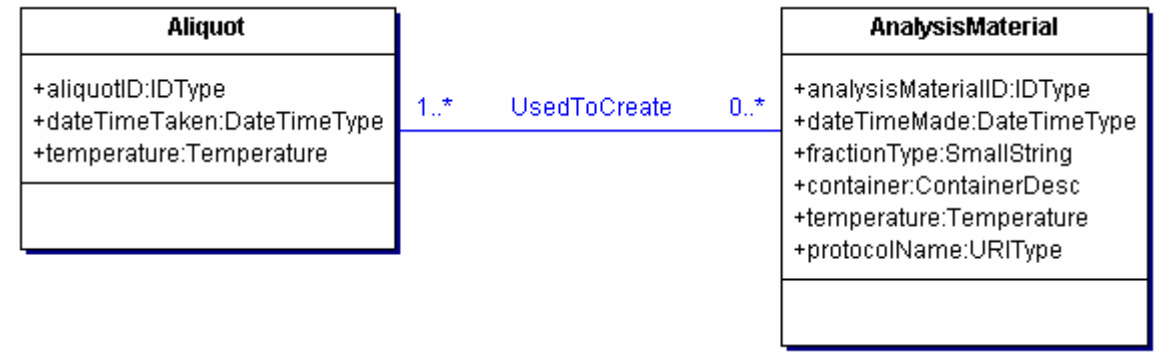
Additional requirements and constraints. The following additional requirements and constraints exist for the Sample entity attributes:

- The five letter entity identifier that makes up part of sampleID should be "sampl".
- The value for dateTimeStored should follow the latest dateTime value for a collection event associated with any of the explants that constitute the sample.

4.6. Sample Preparation

Class Diagram. The class diagram for the Sample Preparation package is depicted in Figure 7.

Figure 7. The Sample Preparation Component



Entities. The Sample Preparation component contains the following entities:

Table 17. Sample Preparation Component Entities

Name	Description
Aliquot	A division of a sample.
AnalysisMaterial	The result of applying preparation protocols to an aliquot.

Dependencies. The Sample Preparation component is dependent upon the Sample Handling component and the Analysis Specific Sample Preparation component. These dependencies can be described by way of the following relationships between entities:

Table 18. Sample Preparation Component Dependencies

Entities	Multiplicity	Relationship
Sample:Aliquot	1..1:0..*	Becomes
PreparationMethod:AnalysisMaterial	0..1:0..*	IsAppliedTo

Attributes. The core data items for the Sample Preparation component are described below. All of these data items are required. Their data types (as given in Figure 7) are described in Appendix A.

Table 19. Sample Preparation Component Aliquot Entity Attributes

Attribute	Description
aliquotID	A unique identifier for the aliquot. (Primary Key)
dateTimeTaken	The date and time at which the aliquot was taken from the sample.
temperature	The temperature at which the aliquot was held prior to preparation.

Additional requirements and constraints. The following additional requirements and constraints exist for the Aliquot entity attributes:

- The five letter entity identifier that makes up part of the aliquotID should be "aliqt".
- The value for dateTimeTaken should follow the value for dateTimeStored for the associated sample.

Table 20. Sample Preparation Component AnalysisMaterial Entity Attributes

Attribute	Description
analysisMaterialID	A unique identifier for the item of analysis material. (Primary Key)
dateTimeMade	The date and time at which the analysis material was produced from the aliquot.
fractionType	The type of the fraction used to create the analysis material following metabolite extraction and fractionation.
container	The nature of the container in which the analysis material is held.

Attribute	Description
temperature	The temperature at which the analysis material is held prior to instrumental analysis.
protocolName	A reference to the protocol followed to prepare the aliquot for instrumental analysis.

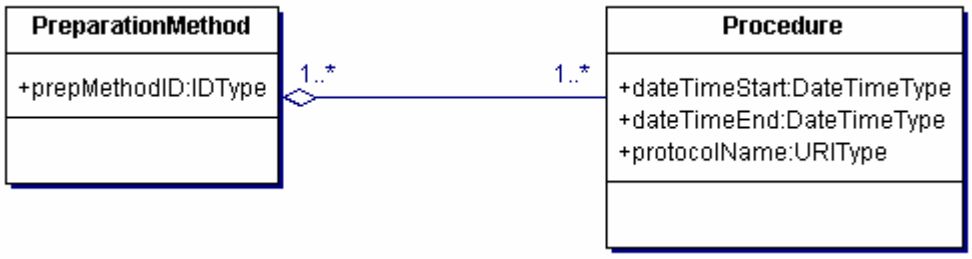
Additional requirements and constraints. The following additional requirements and constraints exist for the AnalysisMaterial entity attributes:

- The five letter entity identifier that makes up part of analysisMaterialID should be "anamt".
- The value for dateTimeMade should follow the values for dateTimeTaken for all associated aliquots.

4.7. Analysis Specific Sample Preparation

Class Diagram. The class diagram for the Analysis Specific Sample Preparation component is depicted in Figure 8.

Figure 8. The Analysis Specific Sample Preparation Component



Entities. The Analysis Specific Sample Preparation component contains the following entities:

Table 21. Analysis Specific Sample Preparation Component Entities

Name	Description
PreparationMethod	The preparation of an item of analysis material for presentation to a particular analytical instrument.
Procedure	A particular procedure which forms part of a preparation method.

Dependencies. The Analysis Specific Sample Preparation component has no dependencies on other components.

Attributes. The core data items for the Analysis Specific Sample Preparation component are described below. All of these items are required. Their data types (as given in Figure 8) are described in Appendix A.

Table 22. Analysis Specific Sample Preparation Component PreparationMethod Entity Attributes

Attribute	Description
prepMethodID	A unique identifier for a preparation

Attribute	Description
	method. (Primary Key)

Additional requirements and constraints. The following additional requirements and constraints exist for the PreparationMethod entity attributes:

- The five letter entity identifier that makes up part of prepMethodID should be "pmeth".

Table 23. Analysis Specific Sample Preparation Component Procedure Entity Attributes

Attribute	Description
dateTimeStart	The date and time at which the procedure was started. (Partial Primary Key)
dateTimeEnd	The date and time at which the procedure ended. (Partial Primary Key)
protocolName	A reference to the protocol followed for this procedure. (Partial Primary Key)

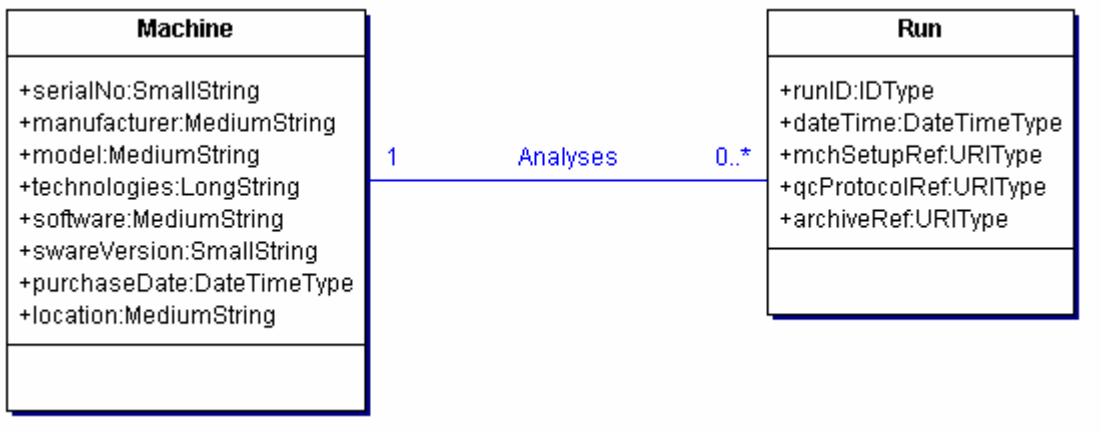
Additional requirements and constraints. The following additional requirements and constraints exist for the Procedure entity attributes:

- The value for dateTimeEnd should preceed the value entered for dateTimeMade for all associated AnalysisMaterial.
- The value for dateTimeEnd should follow the value for dateTimeStart.
- The time periods described by dateTimeStart and dateTimeEnd for all procedures associated with the same preparation method should not overlap.

4.8. Instrumental Analysis

Class Diagram. The class diagram for the Instrumental Analysis package is depecited in Figure 9.

Figure 9. The Instrumental Analysis Component



Entities. The Instrumental Analysis component contains the following entities:

Table 24. Instrumental Analysis Component Entities

Name	Description
Machine	An analytical instrument.

Name	Description
Run	The analysis of an item of analysis material on an analytical instrument.

Dependencies. The Instrumental Analysis component is dependent upon the Admin component and the Sample Preparation component. These dependencies can be described by way of the following relationships between entities:

Table 25. Instrumental Analysis Component Dependencies

Entities	Multiplicity	Relationship
User:Run	1..1:0..*	Operates
AnalysisMaterial:Run	1..1:0..*	IsAnalysedBy

Attributes. The core data items for the Instrumental Analysis component are described below. All of these data items are required. Their data types (as given in Figure 9) are described in Appendix A.

Table 26. Instrumental Analysis Component Machine Entity Attributes

Attribute	Description
serialNo	The serial number of the analytical instrument. (Primary Key)
manufacturer	The name of the manufacturer of the analytical instrument.
model	The model of the analytical instrument.
technologies	The technologies that the instrument employs, e.g. Agilent 7683 Series Injector, Agilent 6890 Series GC System.
software	The name of the software that the instrument employs.
swareVersion	The version number of the software employed by the instrument.
purchaseDate	The date on which the instrument was acquired.
location	The geographical location of the instrument.

Additional requirements and constraints. The following additional requirements and constraints exist for the Machine entity attributes:

- The value for purchaseDate should preceed the values for dateTime in all associated runs.

Table 27. Instrumental Analysis Component Run Entity Attributes

Attribute	Description
runID	A unique identifier for the run. (Primary Key)
dateTime	The date and time that the run took place.

Attribute	Description
mchSetupRef	A reference to the documented setup of the analytical instrument for the run.
qcProtocolRef	A reference to the quality control protocol for the analytical instrument that was being followed at the time of the run.
archiveRef	A reference to an archive copy of the raw output from the analytical instrument for the run.

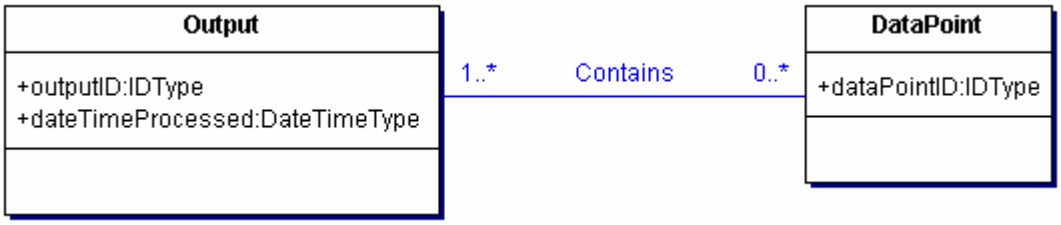
Additional requirements and constraints. The following additional requirements and constraints exist for the Run entity attributes:

- The five letter entity identifier that makes up part of runID should be "runid"
- The value for dateTime should follow the values for dateTimeMade for all associated AnalysisMaterial.

4.9. Metabolome Estimate

Class Diagram. The class diagram for the Metabolome Estimate package is depicted in Figure 10.

Figure 10. The Metabolome Estimate Component



Entities. The Metabolome Estimate component contains the following entities:

Table 28. Metabolome Estimate Component Entities

Name	Description
Output	A set of processed results.
DataPoint	A data point in a set of processed results.

Dependencies. The Metabolome Estimate component is dependent upon the Admin component and the Instrumental Analysis component. These dependencies can be described by way of the following relationships between classes:

Table 29. Metabolome Estimate Component Dependencies

Entities	Multiplicity	Relationship
User:Output	1..1:0..*	Processes
Run:Output	1..1:0..*	Produces

Attributes. The core data items for the Metabolome Estimate component are described below. All of these data items are required. Their data types (as given in Figure 10) are described in Appendix A.

Table 30. Metabolome Estimate Component Output Entity Attributes

Attribute	Description
outputID	A unique identifier for the dataset. (Primary Key)
dataTimeProcessed	The date and time at which the dataset was produced.

Additional requirements and constraints. The following additional requirements and constraints exist for the Output entity attributes:

- The five letter entity identifier that makes up part of ourputID should be "dtast".
- The value of dateTimeProcessed must follow the values for dateTime in all associated runs.

Table 31. Metabolome Estimate Component DataPoint Entity Attributes

Attribute	Description
dataPointID	A unique identifier for the data point. (Primary Key)

Additional requirements and constraints. The following additional requirements and constraints exist for the DataPoint entity attributes:

- The five letter entity identifier that makes up part of dataPointIS should be "datap".

4. Implementation Notes

Component/Entity Naming Conventions. In situations in which the implementation environment invalidates the given name for a component or entity as described in this document, it is suggested that a renaming strategy be applied to all given names for implementation in that environment. This could be achieved by way of a standard suffix or prefix to the given name.

Experiment Attributes. It is suggested that, if an experiment naming policy is not already in effect at an institution for which ArMet is implemented, that a suitable naming strategy is defined and used to constrain the contents of experimentName.

Referential Integrity. We suggest that consideration of referential integrity may be performed at three levels depending on the nature of the implementation. These levels are as follows:

- **Entity Level.** Checks for universal integrity, i.e. uniqueness of primary keys.
- **Component Level.** Checks for foreign keys resulting from internal component relationships.
- **Architecture Level.** Checks for foreign keys resulting from dependencies between components.

A. ArMet Data Types

Table A-1. ArMet Data Types

Name	Format and Size	Allowable Values
UserIDType	Non-empty sequence of up to 20 Unicode characters	UserIDs for valid users of an ArMet implementation
SmallString	Non-empty sequence of up to 50 Unicode characters	
MediumString	Non-empty sequence	

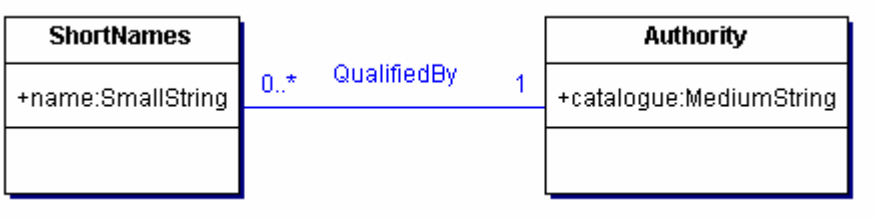
Name	Format and Size	Allowable Values
	of up to 100 Unicode characters	
URIType	Non-empty sequence of up to 200 US-ASCII characters conforming to the Network Working Group RFC 2396	
LongString	Non-empty sequence of up to 200 Unicode characters	
IDType	Non-empty sequence of up to 43 Unicode characters: A five letter entity identifier followed by a 38 digit unique number	
DateTimeType	A date and time value that conforms to ISO 8601	
FloatType	A number that conforms to the W3C Recommendation for XML Schema Datatypes lexical representation for floating point numbers: http://www.w3.org/TR/2001/REC-xmlschema-2-20010502/	
Temperature	See FloatType above	Values in the range -273.00 to +999.00 degrees centigrade

Complex Types. The following types are complex and are described using class diagrams:

Figure A-1. SpeciesNames

Allowable Values. Values from documented catalogues of latin names for species.

Figure A-2. ShortNames

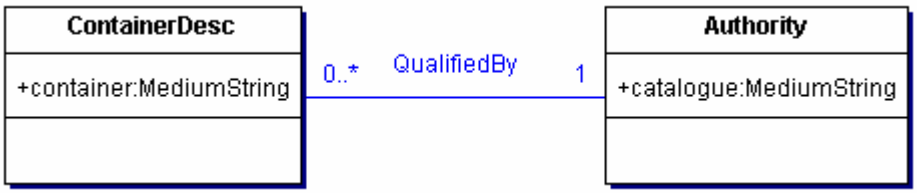


Allowable Values. Values from documented catalogues of line names for species varieties.

Figure A-3. DevType

Allowable Values. Values from documented scales of development stages.

Figure A-4. ContainerDesc



Allowable Values. Values from documented ontologies or controlled vocabularies of container types.