Aberystwyth University

An ontology for a Robot Scientist

Soldatova, Larisa; Clare, Amanda; Sparkes, Andrew; King, Ross Donald

An ontology for a Robot Scientist

Larisa N. Soldatova*, Amanda Clare, Andrew Sparkes and Ross D. King

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

ABSTRACT

Motivation: A Robot Scientist is a physically implemented robotic system that can automatically carry out cycles of scientific experimentation. We are commissioning a new Robot Scientist designed to investigate gene function in *S. cerevisiae*. This Robot Scientist will be capable of initiating >1,000 experiments, and making >200,000 observations a day. Robot Scientists provide a unique test bed for the development of methodologies for the curation and annotation of scientific experiments: because the experiments are conceived and executed automatically by computer, it is possible to completely capture and digitally curate all aspects of the scientific process. This new ability brings with it significant technical challenges. To meet these we apply an ontology driven approach to the representation of all the Robot Scientist's data and metadata.

Results: We demonstrate the utility of developing an ontology for our new Robot Scientist. This ontology is based on a general ontology of experiments. The ontology aids the curation and annotating of the experimental data and metadata, and the equipment metadata, and supports the design of database systems to hold the data and metadata.

Availability: EXPO in XML and OWL formats is at: http://sourceforge.net/projects/expo/. All materials about the Robot Scientist project are available at: http://www.aber.ac.uk/compsci/Research/bio/robotsci/.

Contact: lss@aber.ac.uk

1 INTRODUCTION

1.1 Our new Robot Scientist

A Robot Scientist is a physically implemented robotic system that applies techniques from artificial intelligence to carry out cycles of scientific experimentation (King *et al.*, 2004). A Robot Scientist automatically: originates hypotheses to explain observations; devises experiments to test these hypotheses; physically runs the experiments using laboratory robotics; interprets the results; and then repeats the cycle.

The first Robot Scientist was built in Aberystwyth to investigate *S. cerevisiae* gene function using deletion mutants and auxotrophic growth experiments. In our original proof-of-principle work we demonstrated that a Robot Scientist could rediscover biological knowledge concerning gene function in the aromatic amino acid synthesis pathway. Recently, we have demonstrated that the same approach can be extended to the discovery of novel biological knowledge (King *et al.*, 2005).

*To whom correspondence should be addressed.

An important limitation of our Robot Scientist research has been that although all the intellectual steps were automatic, for some experimental steps it was necessary to intervene manually, owing limitations in our robotic equipment. To eliminate this manual intervention we are commissioning a fully automated Robot Scientist (Figures 1 and 2). This new system is designed to automatically execute yeast growth experiments by: selecting frozen yeast strains from a freezer; inoculating these strains into rich medium; then harvesting a defined quantity of cells; inoculating these cells into specified media (base plus added metabolites and/or inhibitors); and finally accurately measuring growth curves by measuring optical density (OD) (King *et al.*, 2005). We believe, after consulting with the laboratory automation industry, that our new Robot Scientist is one of the most complicated laboratory automated systems in any academic laboratory.

In constructing this new Robot Scientist we have taken advantage of the key benefit of automation: its ability to be easily scaled up. The new Robot Scientist is designed to initiate >1,000 new strain/defined growth-medium experiments a day, using a minimum of 50 different yeast strains, with up to 7 metabolites per experiment, and with each experiment lasting up to 3 days (plus an initiation day). Accurate growth curves will be obtained by observing optical density for every experiment every 20 minutes. This will result in >200,000 data measurements a day. In addition, we expect >1,000,000 meta-data measurements each day. These include hypotheses, experimental plans, experimental actions, temperature, humidity, etc.

1.2 Ontologies for curation and annotation of scientific experiments

Robot Scientists provide unsurpassed test beds for the development of methodologies for the curation and annotation of scientific experiments. This is because, as the experiments are conceived and executed automatically by computer, it is possible to completely capture and digitally curate all aspects of the scientific process: the hypotheses, the experimental goals, the results, etc. The use of a Robot Scientist removes the often 'show stopping' sociological problems associated with trying to capture such data from human scientists.

The ability to capture all relevant experimental information brings with it significant technical challenges:

- We require a very detailed and formalised description of all the domains involved in an experiment: experimental design, methods and technologies; experimental object models and
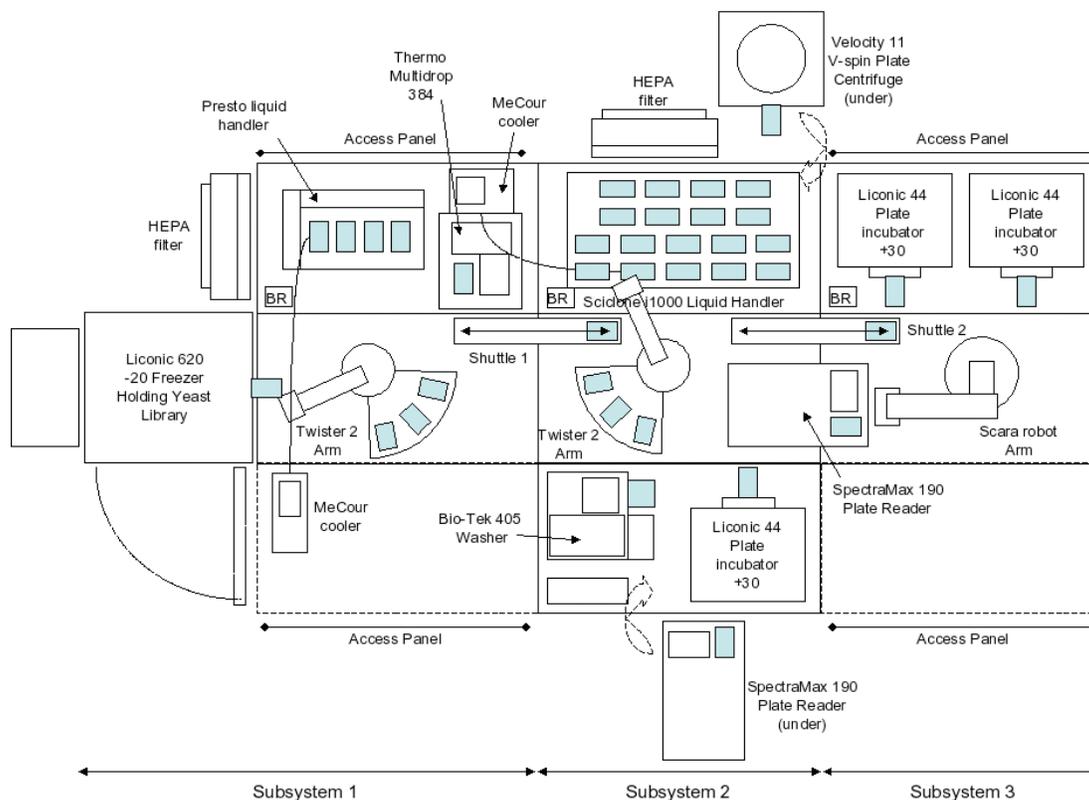
Fig. 1. Plan of our new Robot Scientist.

Fig. 2. Our new Robot Scientist (during assembly, Nov., 2005).

background knowledge; reasoning rules for analysis of the experimental results, etc.

- We need to curate and ensure the integrity of the large amount of data and metadata that the Robot Scientist will produce.
- We wish to make the experimental information as open as possible to both the scientific community and the general public—as part of the mission to improve the public understanding of science.

To meet these challenges we have selected an ontology driven approach to the representation of all the data and metadata relevant to the project. The value of the utilisation of ontologies for the curation and annotation of scientific results is now generally recognised (Bard and Rhee, 2004). The use of ontologies make scientific knowledge more explicit, helps detect errors, enables the sharing and reuse of common knowledge, removes redundancies in domain-specific ontologies, and promotes the interchange and reliability of experimental methods and conclusions.

Bioinformatics has led the way in the application of ontologies to the curation and annotation of experimental data (Brazma *et al.*, 2001). Probably the best known application of ontologies to describing experiments is that developed by the Microarray Gene Expression Society (MGED) (Stoeckert *et al.*, 2002). The MGED Ontology (MO) is designed to provide descriptors required by MIAME (Minimum Information About a Microarray Experiment) standard for capturing core information about microarray experiments. MO aims to provide a conceptual structure for microarray experiment descriptions and annotation. Similar approaches have been made in proteomics (http://psidev.sourceforge.net/ontology/), metabolomics (Jenkins *et al.*, 2004) and anatomy (Ryn and Sternberg, 2003).

Unfortunately, the existing ontologies for experiments representation are not suitable for extension to a Robot Scientist (Soldatova and King, 2005). They are highly human-oriented, and they do not contain concepts about general principles for organising and execution of experiments and analysis of the results. In

addition, no ontology is yet available for microbiological experiments, the domain of the robot scientist experiments.

We have therefore applied our generic ontology of scientific experiments EXPO (Soldatova, 2005) to our Robot Scientist and formed the instantiation EXPO-RS. The goals of this ontology are as follows:

- To formalise the concepts involved in Robot Scientist experiments, and to identify what metadata are essential for the experiment's description and repeatability.

- To provide a controlled vocabulary for all the participants of the project. This includes specialists from different scientific areas (and the general public).

- To organise all the information and knowledge about the Robot Scientist project into different meta-levels. This ensures a clear structure, allows maintenance and updating of the knowledge, and enables coordination of multiple tasks: planning of an experiment; execution of an experiment; access to the results; technical support of the robot, etc.

- To design a database for the storage of experimental data and track experiment execution.

In section 2 we describe a generic ontology of experiments as a method for representation of the information about the Robot Scientist project. Section 3 presents three example applications of the ontology for the Robot Scientist description, namely: its metadata, representation of the data about the experimental equipment and the data base model for storing information about Robot Scientist experiments. Section 4 is devoted to discussion of problems of the data representation for a robot and new challenges.

2 GENERIC ONTOLOGICAL DESCRIPTION OF EXPERIMENTS

We used the generic ontology of scientific experiments EXPO as a method to represent the metadata and data of the Robot Scientist experiments (Soldatova and King, 2006). EXPO provides a clear structured framework for a consistent and shareable description of experiments for both humans and computer systems. EXPO formalises the generic concepts of experimental design, methodology, experimental objects, subjects, equipment, experimental protocols and actions, observations and results representation. EXPO is expressed in the W3C standard ontology language OWL-DL (www.w3.org/TR/owl-guide/). EXPO contains 200 classes and it is available at http://sourceforge.net/projects/expo/.

In defining an ontology, we follow the definition given by Barry Smith[1]: An ontology is a representation of some pre-existing domain of reality which: (1) reflects the properties of the objects within its domain in such a way that there obtains a systematic correlation between reality and the representation itself; (2) is intelligible to a domain expert; (3) is formalised in a way that allows it to support automatic information processing.

To build up the Robot Scientist's ontology EXPO-RS we use the following structure elements:

- **A concept X** (=**class**). 'X is a class if and only if (iff) each element x of X satisfies the intrinsic property of X. The intrinsic

---

property of a thing is a property which is essential to the thing and it looses its identity when the property changes' (Mizoguchi, 2004).

- **An instance x**, an element of the class X.

- **Is-a relation**. '<class A is-a class B> relation holds between classes if and only if (iff) every instance of the class A is also an instance of the class B' (Mizoguchi, 2004). In order to provide a simple hierarchical structure, the concepts are assumed to be disjoint.

- **Instance-of relation**. If and only if (iff) the definition above holds then the relation <x instance-of X> is true.

- **Attribute-of (a/o) relation** is used for describing properties of the concept. It can be considered as a predicate `attribute (Concept, Property)`. This relation can have a fixed cardinality or a range $0, \ldots, n$, where n is a natural number; minimum cardinality 0 means that some of instances of the class might not have this property, i.e. the property is not intrinsic, but still important for the class description as a whole.

- **Part-of relation (p/o)** is used for describing partronomic relations between concepts. For simplicity's sake and because it is not essential for the selected domain, we do not distinguish the different types of whole-part relations (Guarino, 1998). The above comments about cardinality are also true for part-of relations.

All concepts of the Robot Scientist project are defined as subclasses of the following top concepts:

(1) Physical **object**, i.e. experimental equipment.

(2) **Process**, such as an execution of experiment, interpreting the results, experimental actions.

(3) **Proposition**: tasks of experiments, experimental goals, hypotheses, experimental design strategy, models, standards.

(4) **Substrate** for representing time points and intervals, measurement units and locations.

(5) **Role**, for instance functional role, or subject, object role.

The role concept is particularly important for the Robot Scientist because the robot can play different roles in the same experiment:

- The robot is the *object* of an experiment when we study the automation of science. The experimental domain in this case is Artificial Intelligence and Robotics.

- The robot is the *subject* of the experiment when we employ the robot to discover new knowledge in a scientific domain. In this article we concentrate on the description of robot-subject experiments.

EXPO-RS is built as an extension of EXPO by adding the specifics of the Robot Scientist project to the classes and instances.

3 APPLICATIONS OF AN ONTOLOGY FOR THE ROBOT SCIENTIST

3.1 Metadata

We illustrate in Figures 3 and 4 an example of a Robot Scientist experiment annotated using EXPO-RS (King *et al.*, 2005). In

```
<scientific experiment>:
    <admin. info about experiment>:
    <title>:                          Robot scientist
    <ID>:                             exp200401113-0001
<classification by domain>:
    <domain of experiment>:
    <DDC(Dewey) classification>:      576 Microbiology
<research hypothesis>:
<representation style>:               <text>
    <linguistic expression>:          <natural language>:
                                      Knocked out gene named ``yer152c'' (= met8) has the function named
                                      ``2-aminoadipate:2-oxoglutarate aminotransferase'' (E.C.2.6.1.39)
    <linguistic expression>:          <artificial language>:
                                      encodes(yer152c, '2-aminoadipate:2-oxoglutarate aminotransferase')
<null hypothesis>:
    <linguistic expression>:          <artificial language>:
                                      ~ encodes(yer152c, '2-aminoadipate:2-oxoglutarate aminotransferase')
<alternative hypothesis>:
    <linguistic expression>:          <natural language>:
                                      <time effect>: maturation effect (incubator too cold)
<alternative hypothesis>:
    <linguistic expression>:          <natural language>:
                                      <object effect>: no entry of metabolite into the cells
<alternative hypothesis>:
    <linguistic expression>:          <natural language>:
                                      <object effect>: cross contamination
<domain model>:                       <representation style>: <text>
    <linguistic expression>:          <artificial language>: Prolog
                                      A logical model of yeast metabolism
    <reference>:                      Whelan, K.E. & King, R.D. (2005) Using a logical model to predict
                                      the growth behaviour of yeast cell cultures. Department of Computer
                                      Science Report, University of Wales, Aberystwyth. UWA-DCS-05-045.
<experimental design>:
    <subject>:                        The Robot Scientist
    <object>:                         S. cerevisiae
<experimental model>:
    <factor>:                         Strain - 2 strains: wild [Mat A, by4741] and its yer152c knockout
    <factor>:                         addition or not of metabolite 2-aminoadipate:2-oxoglutarate
                                      aminotransferase
    <model assumption>:               stationarity

..................................................
<experimental conclusion>:            <representation style>: <text>
    <linguistic expression>:          <natural language>:
                                      The yer152c knockout strain has a quite different growth profile to
                                      the wild type. This is consistent with yer152c encoding a
                                      2-aminoadipate:2-oxoglutarate aminotransferase. We hypothesize that
                                      yer152c is the missing 2-aminoadipate:2-oxoglutarate
                                      aminotransferase II.
```

Fig. 3. EXPO-RS formalisation of a Robot Scientist experiment in a text format (a fragment).

Figure 3 (and further in the text) the terms in angled brackets are from EXPO-RS. Figure 3 shows the corresponding fragment of EXPO-RS in a text format and Figure 4 in a graphic format (Kozaki *et al.*, 2002).

The goal of the illustrated experiment is to investigate the function of the gene named 'YER152c'. This gene is currently classified by SGD/GO as 'Uncharacterized', and by MIPS as 'Unclassified'. In previous work on predicting gene function we predicted the gene to be involved in 'metabolism' with estimated >80% accuracy (Clare and King, 2003).

The Robot Scientist used its background bioinformatics knowledge in its internal databases to abduce the hypothesis that YER152c encodes the enzyme 2-aminoadipate: 2-oxoglutarate aminotransferase. This is formally encoded in the Prolog fact 'encodes (yer152c, '2-aminoadipate: 2-oxoglutarate aminotransferase')'. Given this abduction, and its general model of yeast metabolism, the Robot Scientist deduced that the removal of this gene would produce a strain with reduced growth (a bradytrophic mutant) or no growth (an auxotrophic mutant); and that addition of the metabolite L-2-aminoadipate to the standard defined growth medium would restore growth. Analysis of the experimental results provided evidence that was consistent with YER152c encoding the missing 2-aminoadipate: 2-oxoglutarate aminotransferase II (N.B. it is a known iso-enzyme: (Masuda and Ogur, 1969)).

The application of an ontology to this experiment demonstrates its value in providing the structure for annotating and curating our Robot Scientist's experimental information. Note in particular, the use of the ontology made explicit: the analysis of alternative hypotheses, assumptions about the domain model and possible factors that could affect the experimental results. Finally, as EXPO is a general ontology of scientific experiments, its application provides the framework to link the Robot Scientist's data and metadata to other scientific data and metadata.

3.2 Description of experimental equipment

Our new Robot Scientist's laboratory automation hardware is extremely complicated and comes supplied with substantial amounts of technical description. Application of an ontology helps to define which of the equipment characteristics are most important to describe to ensure experimental reproducibility.

Fig. 4. EXPO-RS representation of a robot scientist experiment (a fragment).
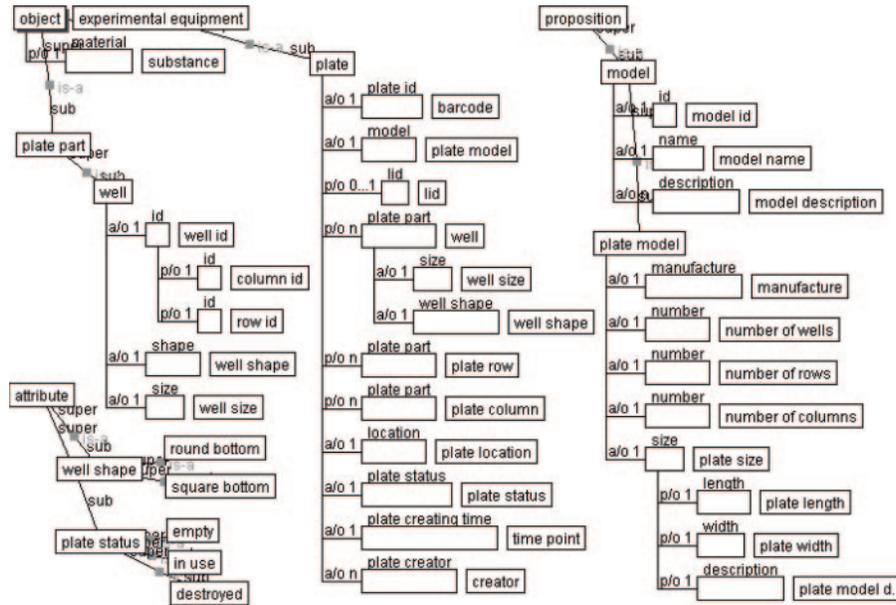

Fig. 5. EXPO-RS representation of the experimental equipment (a fragment for a plate).

A description of the functionality of the equipment highlights the requirement for collection of metadata from the equipment. For example: if the equipment can do an action *A*, do we need to make sure whether or not *A* happened is recorded in our data records; if part of the equipment is replaced due to failure, does the new equipment satisfy the functionality that the old equipment provided, and what are the differences? An ontological description

of this functionality gives us a systematic framework for making decisions about the metadata we need to record, and a framework for comparing metadata collected from differing pieces of equipment.

In EXPO-RS each piece of laboratory equipment is defined through 'physical' object. For example, a well is defined as <plate part> (see Figure 5). As a well cannot exist separately
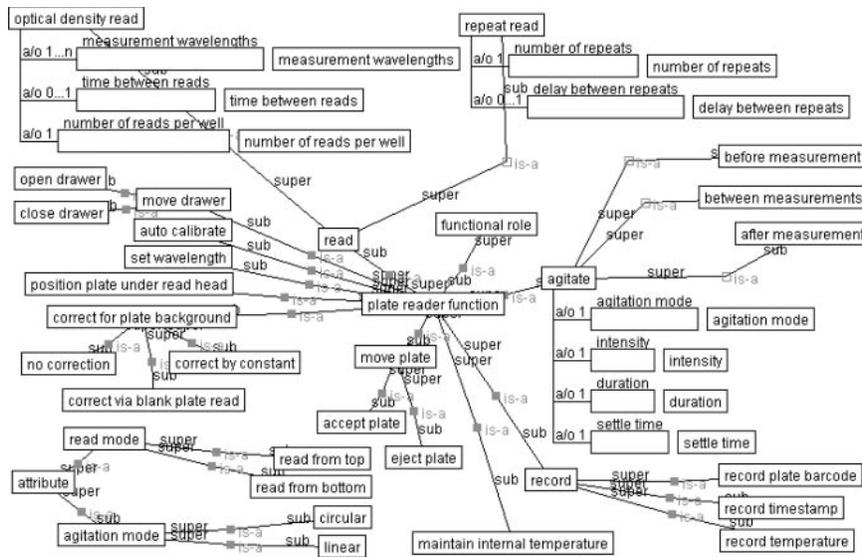
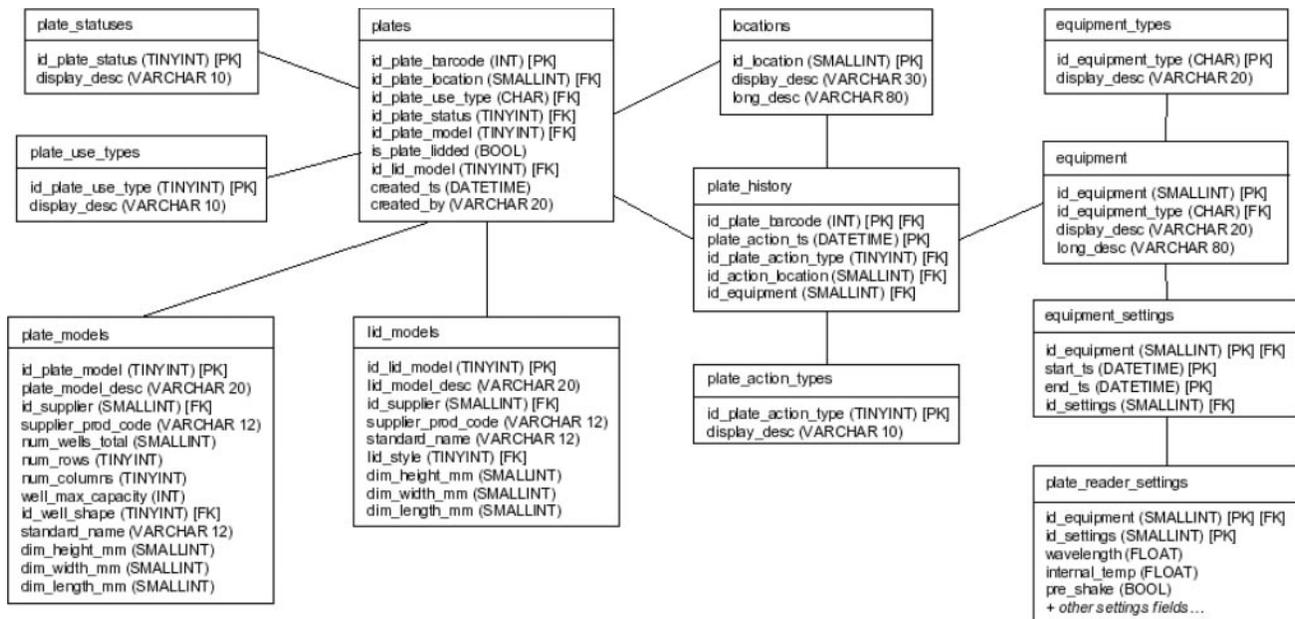**Fig. 6.** EXPO-RS representation of a plate reader functions.

Fig. 7. Data base model for the Robot Scientist (a fragment), where PK is a primary key and FK is a foreign key.

from a plate it cannot be a single object. A representation of a well is essential for representing the Robot Scientist's experimental observations, because optical density is measured in each individual well, and stored by well. The ontology describes a concept <well> with its important characteristics: identification number <well id> (from plate <column id> and plate <row id>); <well shape> that can be <round bottom> or <square bottom>, and <well size>. Note that the attributes <well shape> and <well size> are also used for plate descriptions. The reason for this is that no plate can have wells of differing size and shape. Plates for the pregrowth stage of the experiment will have <round bottom> <well shape>

for better centrifugation separation, while those used in the freezer and in the growth phase will have <square bottom> wells.

Administrative information about the equipment, contact details of suppliers and models information are represented as propositions. Each <model> is characterised by its <id>, <name> and has <model description>. The latter can have different <representations> (not shown) on different representation media such as electronic e.g. a CD or paper e.g. a book. <Plate model> inherits properties of <model> concept and additionally has <plate size>, <number of wells> properties, etc. These attributes are not essential in describing a plate as a piece of experimental equipment or for experiment

representation; thus they characterise a particular plate model and are stored separately. A record of the <plate model> and <manufacturer>, and the same for the plate lid (not shown in this fragment) ensures that experimental variation due to readings of different types of plate may be noticed.

The ontology also contains a description of the equipment functionality. We illustrate the application of this ontology by showing a fragment of the functionality of a plate reader. The fragment in Figure 6 shows the main functions of a Robot Scientist plate reader, that is to perform an <optical density read>, to <agitate> the plate, to <record> information such as the plate barcode, timestamp, and temperature, and to allow the correct parameters of the read and agitation to be set. The plate readers currently used in the Robot Scientist are two SpectraMax 190s. This information would be recorded as the model name under the part of the ontology that describes the equipment. Describing these readers in terms of the plate reader functionality part of the ontology enforces a record of the specifics of our laboratory setup. The <number of readings per well> is one, and there is one <measurement wavelength> at 595nm. The reader does <maintain internal temperature> at 30 degrees C. Usually the reader does not <agitate> the plate, as the plates are continually agitated while they are in the incubators but there is one occasion on which the reader must <agitate> <before measurement> for a <duration> of 30 seconds. This is to resuspend the yeast after centrifugation. This particular model of plate reader does not inform us what the <intensity> or <agitation mode> are. All this metadata is to be recorded in our database.

The next section describes the use of the ontology in the design of this database.

3.3   Use of the ontology for the design of a data model

As described above, a Robot Scientist will generate a very large amount of data and metadata. To ensure the integrity of this data, and to provide for its easy access, we will store all the data and metadata in a relational database.

The principal application of the ontology to database design was as an aid to identifying objects and events that needed to be recorded in the database. This was of key importance, as the primary aim when creating a good relational database design is to model the real world system as closely as possible. You first identify the objects and events that you want the database to represent: creating a structured ontology of your system is a good way of doing this. You then define the tables and all the relevant fields that they should contain, and finally describe how they are all related.

The ontology also helped with naming both tables and columns, with defining relationships between various data, and as a verification that the database design had incorporated all of the data useful to the project.

The fragment of the database design shown in Figure 7 handles the data records of individual 96-well plates; what model of plate and lid it is, what use it is being put to, what actions have happened to it during its lifetime within the Robot Scientist project, and the details of the robotic equipment that have been used to handle it. For each piece of equipment (e.g. a plate reader) it stores what settings were used and over what timeframe. This allows you to retrieve exactly what settings were used on any piece of equipment that interacted with any particular plate at any time in the history of the project.

To explain the main 'plates' table columns in more detail:

- **id_plate_barcode**: Each physical 96-well plate has a unique 8-digit <barcode> label attached to it for tracking purposes. There are three barcode readers on the Robot Scientist, one for each of the three subsystems (see Fig. 1). The plate is scanned once in the first subsystem to *create* it, and again on entry into subsystems two and three to check its identity before it is worked on. For example, 00012345.

- **id_plate_location**: Each physical position on the Robot Scientist where a plate can be placed or moved to has a unique <location> number, with all valid locations stored on the separate 'locations' table. For example the plate reader in subsystem three is location 3300.

- **id_plate_use_type**: This is a reference to the <plate usage> for the specific plate. These are held on the 'plate_use_types' table. There are currently three uses a plate could be put to; as a <yeast strain library plate>, as a <yeast pregrowth plate>, and as an <experiment nutrient cocktail plate>.

- **id_plate_status**: Each plate has a <status> associated with it to record its current condition. Generally a plate will initially start off in an <empty> state, then become <in use>, and then when it is finished with and disposed of it becomes <destroyed>. This allows us to quickly identify which plates are active and which are historical.

- **id_plate_model**: This is a reference to the <model> of plate, we use different models for different parts of the system; for example the yeast library plates are larger to accommodate greater volumes in deeper wells, whilst the experiment cocktail plates are made of clear polystyrene and have flat-bottomed wells to allow optical readings to be taken. Similar plates may also be made by different manufacturers so we need to record this. The various models of plate are stored on the 'plate_models' table which in turn is linked to supplier information (not shown).

- **is_plate_lidded**: A Boolean flag to indicate whether the plate has a <lid> or not.

- **id_lid_model**: This is a reference to the <lid model>. For example a lid may be flat or it may have ridges to reduce evaporation from wells. The various models of lid are stored on the 'lid_models' table.

- **created_ts**: This field is used to store the <timestamp> (time and date) of when the plate was created. In database terms this refers to the first time its unique barcode was scanned, normally when a robot arm has first taken it from a consumables plate stack for use.

- **created_by**: This field is used to store who or what created the plate. If the plate was manually created and introduced to the system (e.g. a yeast library plate) this field will contain the name of the person who set it up. Otherwise it will contain a name related to where on the Robot Scientist it was created.

As in the application of EXPO to curating and annotating experimental metadata, and the curation and annotation of metadata on experimental equipment, the application of a general experimental ontology to database design allows data and metadata to be compared and shared between experiments and laboratories.

4   DISCUSSIONS AND CONCLUSION

A Robot Scientist enables us to capture an unprecedented amount of information about scientific experiments. For the first time it is possible to completely capture and digitally curate all aspects of the scientific process. This presents us both with unique opportunities and challenges. The opportunity is the ability for the first time to record and fully understand how and why a particular experiment was conceived and executed, and to remove all subjectivity in experimental actions. This enables all aspects of experimentation, including hypothesis formation and testing, to be fully repeatable.

The great technical challenge is how to capture and digitally curate all this information. We argue that formation of a Robot Scientist ontology is a key step in meeting this challenge. We have used such an ontology to curate and annotate the experimental data and metadata and the equipment metadata, and to help design the associated database systems. As our ontology is linked to a general ontology of scientific experiments (EXPO) all the data and metadata captured can be shared with other experiments. We envisage our ontology as a start point for further community efforts in developing a general ontology for fully automated laboratories.

We believe that this increased ability to record and curate all aspects of scientific experiments will have important ramifications for scientific publishing. As in the e-Science 'vision' it will be increasingly easy to link papers to all the relevant data and metadata, ensuring full repeatability. In this task we believe that natural language will be required less and less to describe experiments. This is to be welcomed as natural language is notorious for its imprecision and ambiguity. Its use is also a great hindrance when using computers to store and analyse data—hence text-mining. We therefore argue that the content of scientific papers should increasingly be expressed in formal languages with ontological foundations.

ACKNOWLEDGEMENTS

REFERENCES

Bard,J. and Rhee,S. (2004) New ontologies in biology: design, applications and future challenges. *Nature Reviews. Genetics*, **5/3**, 213–222.

Brazma,A., Hingamp,P., Quackenbush,J., Sherlock,G. *et al.* (2001) Minimum information about a microarray experiment (MIAME):toward standards for microarray data. *Nature Genetics*, **29**, 365–371.

Clare,A. and King,R.D. (2003) Predicting gene function in *Saccharomyces cerevisiae*. In *Proceedings of the 2nd European Conference on Computational Biology*.

Guarino,N. (1998) Some ontological principles for designing upper level lexical resources. In *First International Conference on Language Resources and Evaluation (eds. Rubio, Gallardo, Castro, Tejada)*, 527–534.

Jenkins,H., Hardy,N., Beckmann,M., Draper,J. *et al.* (2004) A proposed framework for the description of plant metabolomics experiments and their results. *Nature Biotechnology*, **22** (12), 1601–1606.

King,R.D., Whelan,K., Jones,M., Reiser,P. and Bryant, C. (2004) Functional genomics hypothesis generation by a robot scientist. *Nature*, **427, no 6971**, 247–252.

King,R., Young,M., Clare,A., Whelan,K. and Rowland,J. (2005) The robot scientist project. In *Springer Lecture Notes in Computer Science 3735 (eds. A.G. Hoffmann, Motada, H., Scheffer, T.)*, pages 16–25.

Kozaki,K., Kitamura,Y., Ikeda,M. and Mizoguchi,R. (2002) Hozo: An environment for building/using ontologies based on a fundamental consideration of ''role'' and ''relationship''. In *Knowledge Engineering and Knowledge Management*, 213–218.

Masuda,M. and Ogur,M. (1969) Enzymatic and physiological properties of the yeast glutamate-alpha-ketoadipate transaminase. *J. Biol. Chem.*, **244**, 5153–8.

Mizoguchi,R. (2004) Tutorial on ontological engineering, Part 3: Advanced course of ontological engineering. *New Generation Computing*, **22/2**, 193–220.

Ryn,L. and Sternberg,P.W. (2003) Building a cell and anatomy ontology of *Caenorhabditis elegans*. *Comparative and Functional Genomics*, **4**, 121–126.

Soldatova,L. and King,R. (2005) Are the current ontologies used in biology good ontologies? *Nature Biotechnology*, **9/23**, 1096–1098.

Soldatova,L. and King,R. (2006) An ontology of scientific experiments. *Journal of the Royal Society Interface*, (in press).

Soldatova,L. (2005) EXPO. http://sourceforge.net/projects/expo/

Stoeckert,C., Causton,H. and Ball,C. (2002) Microarray databases: standards and ontologies. *Nature Genetics*, **32**, 469–473.