An Ontology for Description of Drug Discovery Investigations

Da Qi[1], Ross D. King[1], Andrew L. Hopkins[2], G. Richard J. Bickerton[2], Larisa N. Soldatova[1*]

[1]Department of Computer Science, Aberystwyth University, UK, SY23 3DB

[2]Division of Biological Chemistry and Drug Discovery, University of Dundee, UK, DD1 5EH

Summary

The paper presents an ontology for the description of Drug Discovery Investigation (DDI). This has been developed through the use of a Robot Scientist "Eve", and in consultation with industry. DDI aims to define the principle entities and the relations in the research and development phase of the drug discovery pipeline. DDI is highly transferable and extendable due to its adherence to accepted standards, and compliance with existing ontology resources. This enables DDI to be integrated with such related ontologies as the Vaccine Ontology, the Advancing Clinico-Genomic Trials on Cancer Master Ontology, etc.
DDI is available at http://purl.org/ddi/wikipedia or http://purl.org/ddi/home .

1 Introduction

The accumulated historical record of experimental data is one of the most valuable intellectual property assets of pharmaceutical companies. However, storing experimental data and protocols in sufficient detail to ensure exact reproducibility has proven difficult. The result is that the extended utility of data or protocols beyond their projects has rarely been demonstrated [1, 2]. The fundamental problem of data exchange and data integration in the pharmaceutical industry is the lack of formalised agreement on what data and metadata of drug discovery experiment should be recorded, and how these data should be unambiguously stored. Recently pharmaceutical companies have begun to explore the possibility of developing, in a pre-competitive way, informatics standards to exchange data within the industry and between industry and academia [3]. In initiatives such as the Pistoia Alliance[1], pharmaceutical companies have begun to define a common workflow with a view to standardising processes and terms in the drug discovery process. In developing these standards the Pistoia Alliance aims to utilise the emergence of semantic based web technologies and service-oriented architectures. This recognition that the future informatics framework for pharmaceutical research will be based on exchangeable semantic terms [3] creates the need for an ontological framework for experimental drug discovery data.

The Harvard Business Review lists the need for a common digital data standard in drug development as one of their 10 breakthrough ideas for 2010: "One change would make a substantial difference: the creation of agreed-upon standards for digitally representing drug assets. The challenge is that every company has its own idiosyncratic (and therefore redundant) means of

*Correspondence:lss@aber.ac.uk

[1] http://pistoiaalliance.sourceforge.net/comms/PistoiaAllianceACSSaltLakeCityMarch2009.pdf

collecting, storing, and exploiting information from development trials, making it difficult to share the hundreds of gigabytes of documents and images among partners." [4]

The use of ontologies is becoming increasingly important in scientific research. One of the most important applications of ontologies is in the standardisation of the annotation of experiments. Ontology development for experiment annotation in transcriptomics, proteomics, and metabolomics is well advanced[2]; although many problems still remain. Biology has led the way in applying ontologies to science, and the utility of ontologies has been clearly demonstrated in several biological domains, e.g. Gene Ontology [5], Metabolomics Standards Initiative [6].

Here we propose an ontology for drug discovery investigations (DDI). The purpose of the DDI ontology is to add value to the information generated in the drug discovery pipeline by making information generated easier to reuse, integrate, curate, retrieve, and reason with. DDI will also support information exchange, as companies often have great difficulty exchanging information on drug discovery (they may be merging, one company is selling the information to the other, etc.), as their databases/data-standards are typically not comparable. DDI will minimize this difficulty by providing a standard way for information to be mapped between databases.

1.1 The drug discovery pipeline

Drug discovery is a complex and long-term scientific investigation. It involves a number of phases that together make up the so called 'drug discovery and development pipeline' (Figure 1). The two main phases are preclinical research, and clinical development. Arguably the division between these is the first testing of an experimental drug in humans. In essence the goal of the preclinical research phase is to discover potential drug candidates that are suitable for clinical trials. It involves target discovery and validation, assay development, and lead generation and optimisation. The goal of the clinical research phase is to understand the safety of the compound in humans, and to confirm the efficacy of the drug.

Figure 1: Drug discovery pipeline (see [7] for alternative pipeline drawing)

Various drug discovery (preclinical) process pipelines can be constructed depending on the strategy that is employed [7]. For example, a forward chemical genetics approach starts with the screening of compounds to identify those which affect a phenotypic assay in a desired manner. In contrast, a reverse chemical genetics approach begins with a molecular target of interest and attempts to discover compounds which modulate that target in a desired way. The standard pipeline process model of drug discovery normally now assumes a reverse chemical genetic approach at its core.

[2] http://obi-ontology.org

DDI aims to support recording of data and metadata for the research and development phase and to be combined with formalisms supporting other phases of the drug discovery pipeline.

1.2 Robot Scientists

A Robot Scientist is a physically implemented computer/robotic system that utilises techniques from artificial intelligence (AI) to execute cycles of scientific experimentation [8]. A Robot Scientist is designed to automatically: originate hypotheses to explain observations, devise experiments to test these hypotheses, physically run the experiments using laboratory robotics, interpret the results, and then repeat the cycle. Robot Scientists have the potential to increase significantly the speed and effectiveness of the scientific discovery process and so reduce its cost [9]. Our Robot Scientist "Adam" is the first to demonstrate the automated discovery of novel scientific knowledge [8].

Our new Robot Scientist "Eve" is "a prototype system to demonstrate the automation of closed-loop learning in drug-screening and design" [10]. Eve's robotic system is capable of moderately highthroughput compound screening (greater than $10,000$ compounds per day) and is designed to be flexible enough such that it can be rapidly re-configured to carry out a number of different biological assays. It is able to automatically switch from mass screening mode to QSAR learning. Therefore with Eve there is no need to wait until all compounds in a compound library are screened to start a QSAR process. DDI has been developed for and being used to support the recording of data and metadata generated by Eve in explicit semantic form. By the end of the Eve project, drug discovery data and metadata will be publicly available at our project website, in the same way how we made available the data and metadata generated by Adam and semantically annotated with an ontology for LABOratory Robot Scientists (LABORS): http://www.aber.ac.uk/en/cs/research/cb/projects/robotscientist/

1.3 Existing related ontologies

An ontology is "a concise and unambiguous description of what principle entities are relevant to an application domain and the relationship between them" [11]. The proposed DDI ontology is orthogonal to existing ontolgies described below and can be integrated with them.

OBO Foundry (Open Biomedical Ontologies) [12] is an ontology library containing a set of orthogonal interoperable reference ontologies in the biomedical domain and provides a set of principles for ontology development[3]. The Basic Formal Ontology (BFO)[4] provides the top-level classes under which OBO Foundry ontologies should build, while the Relation Ontology (RO) [13] provides the relations that should be used. The use of the same top level classes and relations guarantees a full compatibility and interoperability within OBO and supports cross-domain quires and reasoning.

Ontology for Biomedical Investigations (OBI)[5] is an integrated ontology for the description of investigations in the area of biology and medicine. OBI is developed through collaborations

[3] http://www.obofoundry.org/crit.shtml
[4] http://www.ifomis.org/bfo
[5] http://obi-ontology.org

among 19 biomedical communities (i.e. metabolomics, proteomics, toxilogy, etc.), and is a candidate ontology for the OBO Foundry. The Robot Scientist projects joined the OBI Consortium in 2008. The DDI ontology is an application of OBI for the area of drug discovery. DDI is built on our previous work EXPO (a generic ontology of experiments) [14] and LABORS [8]. DDI also uses ontology of information artifacts (IAO)[6], which is a spin-off the OBI project, for the description of information content entities and Phenotypic Quality Ontology (PATO)[7] for the description of qualities.

There are several other related projects. Infectious Disease Ontology (IDO)[8] aims to define entities relevant to both biomedical and clinical aspects of infectious diseases generally. Subdomain specific extensions of the core IDO complete the set providing ontology coverage of entities relevant to specific sub-domains of the infectious disease field, such as specific diseases or specific areas of research. To ensure consistent representation of vaccine knowledge and to support automated reasoning, a community-based effort to develop the Vaccine Ontology (VO) has been initiated[9]. The intention of the Advancing Clinico-Genomic Trials on Cancer (ACGT) Master Ontology (MO)[10] is to represent the domain of cancer research and management in a computationally tractable manner. The Ontology of Clinical Research (OCRe)[11] is a formal ontology for describing human studies.

The rest of the paper is organised as follows: section 2 provides a description of the design principles and key entities of the proposed DDI ontology, section 3 demonstrates applications of DDI. In section 4, conclusions are made and future works are planned.

2 An ontology for the description of drug discovery investigations

DDI aims to define the principle entities and the relations in the research and development phase of the drug discovery pipeline. DDI is designed to be highly transferable and extendable due to its adherence to accepted standards and compliance with existing ontology resources. This enables the integration of DDI with such related ontologies as VO, ACGT, etc. These features of DDI enable it to be developed to cover the whole drug discovery pipeline.

In developing DDI we followed the OBO Foundry principles. We employed BFO, IAO and OBI to define the top level classes and we used relations defined in RO, IAO and OBI. We developed DDI as an application of OBI for drug discovery by extending the corresponding classes. DDI imports terms from Chemical Entities of Biological Interest (ChEBI) [15], e.g. *chebi:molecular entity*. DDI follows the Minimal Information to Reference External Ontology Terms (MIREOT) [16]. DDI is expressed in a W3C standard Web Ontology Language OWL-DL[12]. DDI includes the following main branches (shown in Figure 2):

The class *ddi:chemical entity* describes the principle molecular entities, its parts and chemical

[6] http://code.google.com/p/information-artifact-ontology/
[7] http://obofoundry.org/wiki/index.php/PATO:Main_Page
[8] http://www.infectiousdiseaseontology.org/Home.html
[9] http://www.violinet.org/vaccineontology/
[10] http://www.ifomis.org/wiki/ACGT_Master_Ontology_%28MO%29
[11] http://rctbank.ucsf.edu/home/ocre.html
[12] http://www.w3.org/TR/owl-features/

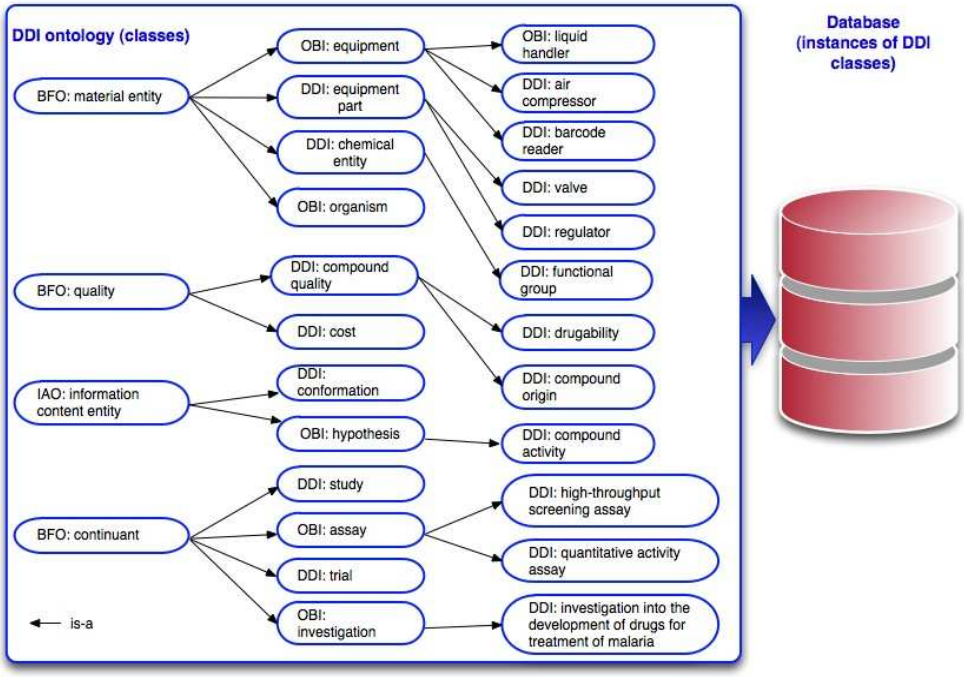

Figure 2: DDI structure

solutions in drug screening and discovery investigations in generic terms. The class *ddi:chemical entity* is a subclass of the class *obi:material entity* which subsumes *object*, *object part*, and *object aggregate*. It additionally defines classes relevant to drug design. For example, the class *ddi:small molecule* (a monomer with weight under 800 Daltons), *ddi:functional group* (a part of a molecule responsible for the characteristic of chemical reactions of such molecule).

The class *bfo:processual entity* describes processes and processes aggregates. The phases in drug discovery pipeline are defined as subclasses of the class *obi:planed process*, e.g. *ddi:assay development phase*; *ddi:hit confirmation phase*. The classes *ddi:drug discovery pipeline*, *ddi:research development phase* are modelled as *bfo:processes aggregate*. Chemical reactions and interactions are subclasses of the class *bfo:process*. DDI applies a five-level approach to describe a specific investigation, which are top to bottom: investigation, study, trial, assay, and replicate, which are modelled as subclasses of the class *bfo:process* (see section 3.1 for more detail).

The class *bfo:quality* contains the entities which describe the characteristics of material entities, such as chemical entities, equipment. DDI defines such qualities as *ddi:compound quality*, *ddi:compound origin* (natural or synthetic), *ddi:drugability* (the likelihood of being able to modulate a target with a small-molecule drug), *ddi:compound library quality* (e.g. diversity). DDI also imports terms form PATO. For example, *pato:length*, *pato:depth*, *pato:width* for the description of equipment; *pato:odor*, *pato:solubility* for the description of chemical entities.

The classes *ddi:equipment* (a material entity that is manufactured by an organisation or person, designed with the intent to perform a specific function or functions) and *ddi:equipment part* are subclasses of the class *obi:processed material* (a material entity that is created or changed during material processing). DDI extends OBI for the description of equipment and equipment parts used by the Computational Biology laboratory at Aberystwyth University, e.g. *ddi:robot*,

ddi:robot arm, *ddi:air compressor*, *ddi:barcode reader*. These classes define generic equipment; a specific equipment can be added as an instance of the corresponding class. For example, Eve uses three different types of liquid handler to conduct its trials. They are instances of the class *obi:liquid handler*. DDI can be easily extended by adding new classes for equipment which is used in other drug screening and discovery laboratories or companies.

The class *bfo:role* for the description of role entities which must be played by some material entity in certain context. For instance, *ddi:compound role* and *obi:drug role* are defined as subclasses of the class *bfo:role*. A chemical entity plays a compound role in most phases of drug discovery pipeline, and plays a drug role once it is approved as a drug. This arrangement allows identity of a material entity to remain unchanged during its lifetime. DDI defines such essential for drug design roles as *ddi:drug target role*, *ddi:inhibitor*, *ddi:hit*, *ddi:lead*, etc. and imports such roles as *chebi:agonist*, *chebi:antagonist*.

The class *iao:information content entity* defines entities that are generically dependent on some artifact and stands in relation of aboutness to some entity. For example, *obi:measurement datum* is a subclass of the class *iao:information content entity* that is the output of an assay. DDI extends these descriptions by adding the classes *ddi:fluorescence polarisation reading*, *ddi:optical density reading* as specified outputs of mass screening assays run by Eve. The class *obi:plan specification* includes parts such as *iao:objective specification*, *iao:action specification* which are subclasses of *iao:information content entity*. DDI adds classes *ddi:objective to find hit-set*, *ddi:objective to find activity* (see Figure 3). DDI also defines such information content entities as *ddi:conformation* (the spatial arrangement of the atoms affording distinction between stereoisomers which can be interconverted by rotations about formally single bonds), *ddi:supply format* (the format of products provided by a supplier) for the description of *ddi:compound supply format* (e.g. powder, liquid).

The DDI assessment against the OBO Foundry principles and other commonly accepted criteria is summarised in the Table 1.

Table 1: A summary of DDI assessment

Principles	**DDI assessment against the principles**
The ontology must be *is a* complete.	All the DDI classes are connected via *is a* relation. There are no “orphan” classes.
The ontology is in a common shared syntax.	DDI is expressed in W3C standard Web Ontology Language OWL-DL.
The ontology includes textual definitions for all terms.	All DDI terms have textual definitions.
The ontology follows an accepted upper level ontology.	DDI uses BFO, OBI and IAO as the top ontologies.
Coverage of the domain	DDI aims to provide descriptors for the research and development phase in the drug discovery pipeline. Currently DDI covers the drug discovery investigations run by the Computational Biology Group at Aberystwyth University. We believe these are typical of the domain.

The ontology uses relations which are unambiguously defined following the pattern of definitions laid down in the OBO Relation Ontology.	DDI uses relations defined in RO, IAO, and OBI. The relations defined in IAO and OBI are candidates for inclusion into RO.
The ontology must be open source.	DDI is open and available on its wiki page. It is designed to be used without any constraint. However, the original source must be acknowledged and it must not be redistributed using the same name and the same identifiers.
The otology possesses a unique identifier space within the OBO Foundry.	The prefix *ddi* is the unique identifier to all DDI terms.
The ontology provider has procedures for identifying distinct successive versions.	DDI is developed under the version control system SVN. Changes in DDI were committed to the SVN repository and were annotated.
The ontology has a clearly specified and clearly delineated content.	DDI is orthogonal to other ontologies already lodged within OBO.
The ontology is well documented.	DDI is documented in its wiki page for distribution. More documentation will be provided for a stage of submission DDI to OBO.
The ontology will be developed collaboratively with other OBO Foundry members.	The DDI team has already started collaboration with the developers of VO. More OBO Foundry members will be invited for collaboration on the next stage of the DDI project.
Multiple inheritance should be dealt with via defined classes.	In DDI, each class has only one superclass. This reduces the potential inconsistency and errors in reasoning processes.
No class can have a single subclass.	Each DDI class has either more than one subclass or none.

3 Applications

DDI provides a framework for describing the knowledge within drug screening and discovery domain and for recording the detailed experimental processes.

3.1 The structure of Eve investigations

DDI allows to explicitly and accurately record metadata about investigations, particularly about a structure of investigations (Figure 3). DDI extends the OBI definition of investigation's structural units. OBI aims to describe the most typical investigations in the area of biomedicine performed by human investigators. The OBI class *obi:investigation* (a planned process that consists of parts: planning, study design execution, documentation and which produce conclusion(s)) defines a biomedical investigations no matter how small or large it is. The class *obi:assay* is used to describe "a planned process with the objective to produce information

about some evaluant". The class *obi:study design execution* is defined as "a planned process that realizes the concretization of a study design".

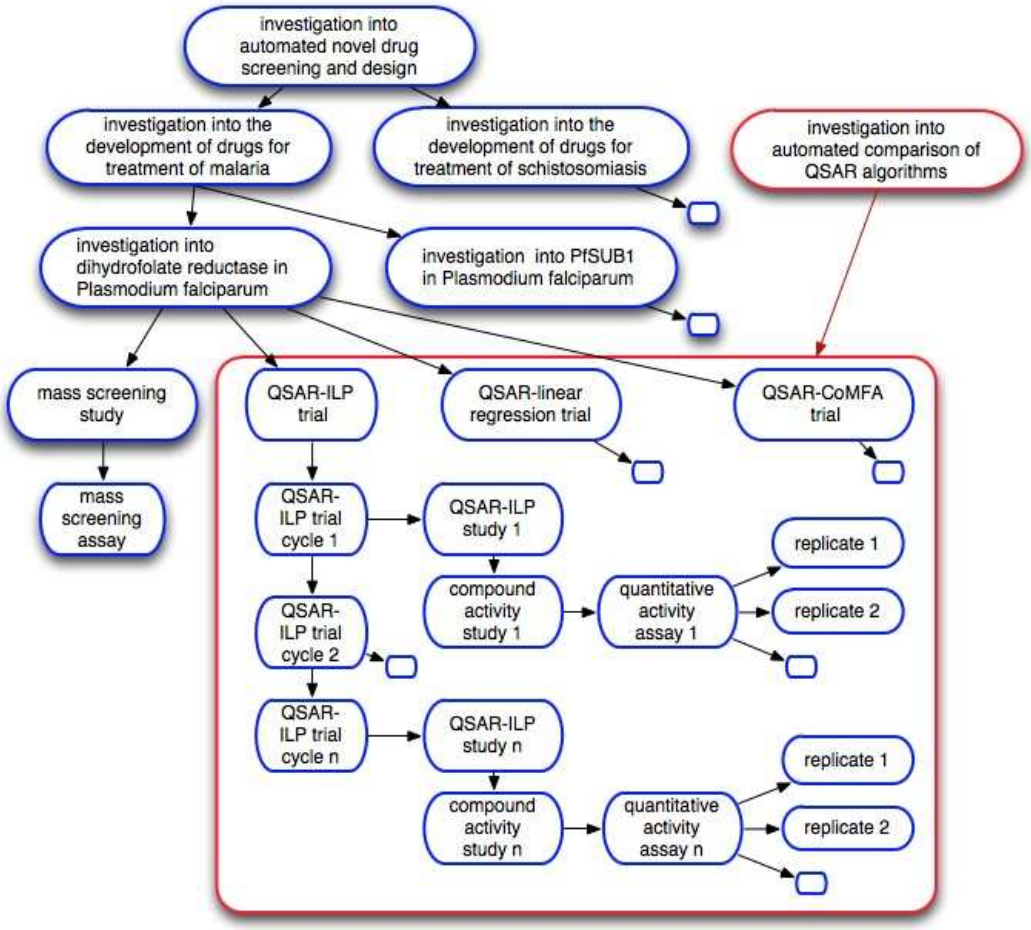

Figure 3: Structure of Eve investigations (blue boxes) and the reuse investigation (red box). The relations between the structural research units are *part of* relations.

It is challenging to comprehensively describe investigations with the use of the only three classes when an investigation has a significantly complex structure. This is especially true for the automated investigations run by Robot Scientists, where thousands of hypotheses are tested in parallel, in cycles, and on different levels of granularity [8]. OBI defines investigations and study design executions in such a way that they cannot have inputs. For example, hypotheses formed in an *obi:hypothesis generating investigation* (an investigation in which data is generated and analyzed with the purpose of generating new hypothesis) cannot be passed to an *obi:hypothesis driven investigation* (an investigation with the goal to test one or more hypothesis) (see also the classes *expo:hypothesis forming investigation* and *expo:hypothesis generating investigation* [14]). To overcome these difficulties, DDI defines structural research units on various levels of granularity. The term *obi:investigation* is reserved for large investigations where metadata such as a leading institution, partner institutions, a project, a PI, a funding body, domains (specified by one or more accepted classification systems), general goals and hypotheses, a time period are recorded. The term *ddi:study* (a planned process which may consists of parts: study design execution, trial, assay, trial cycle, replicate and which produces study results and conclusion(s)) is used for smaller portions of research work performed, where metadata such

as a domain (one from the list of specified domains for the investigation), an investigator, more specific hypotheses, time points are recorded, and it can have input information. Studies are usually parts of a corresponding investigation, and information about a leading institute, funding etc. can be inferred via part of relations. A study can also be separate from an investigation research unit. DDI defines the class *ddi:trial* (a planned process which consists of trial cycles) to represent cyclic portions of the research performed. Eve analyses the results of each cycle in order to design and run the next cycle of research. Currently OBI does not support recording of cyclic research. DDI also defines the class *ddi:replicate* for assays which use the identical study design (e.g. a plate layout). Because the investigations are run by robots, it is possible to accurately repeat assays many times in order to collect required statistics. This approach allows Eve to detect even minor statistically significant differences in responses which would be missed by human observation.

The Figure 3 shows a fragment of the investigation into the development of drugs for treatment of malaria as a part of the investigation into automated novel drug screening and design. The overall goal of the upper-level investigation is to fully automate drug screening and design. The developed technology will be applied to the design of drugs targeting 3^{rd} world diseases, e.g. malaria, and schistosomiasis. Different organisms and different targets will be investigated. The distinction of investigations on these levels is important. In the investigation into automated novel drug screening and design Eve plays a role of a subject of the investigation, and the Computational Biology group at Aberystwyth University is the investigator. We are studying whether a Robot Scientist such as Eve is capable of fully automatic drug design for the specified diseases and organisms. This investigation is from the domain of AI and Robotics, and the hypotheses and the conclusions are formed and expressed in terms of that domain. The investigations into drug screening and design of the selected diseases are run by Eve as the investigator. It is interesting to note that Eve is designed to run investigations, and therefore an investigator is not a role for Eve, but a function. This differs from humans who are not designed to do drug discovery experiments, and for whom an investigator is a role. The recording of the investigations on different levels allows the expression of such differences and the avoidance of logical contradictions.

The first part of investigations into a specified disease in a specified organism is a mass screening study. A compound library is screened in order to find hits – indications of activity of some compounds. Eve makes a decision to stop screening if the number of hits found is sufficient for analysis and prediction of activity by a QSAR trial. Within the Robot Scientist project, different QSAR trials will be run: QSAR-ILP trial, QSAR linear regression trial, QSAR-CoMFA trial, etc. Each cycle of each QSAR trial consists of a computational QSAR compound activity study where the specified input from an assay is analysed and predictions about compound activities are made. The predicted active compounds are not necessarily from the available compound library. The predicted active compounds could be ordered from other commercial compound libraries, or specially synthesised. The predicted active compounds are tested at the next physical quantitative activity assay with many replicates, the results of which are used for the next QSAR study, and so on until Eve makes the decision that a set of leads is found.

The goal of this "intelligent" approach to screening is enable a significant reduction in the size of a compound library, and thereby the cost of drug screening experiments. The initial size of our library is only $\sim 15,000$ compounds. The definition of structural units of investigations is important for the efficient analysis and reuse of the collected experimental data. For example, the results from different QSAR trials can be easily reused for a new investigation into

comparison of QSAR algorithms (see Figure 3 red box).

3.2 Mass screening

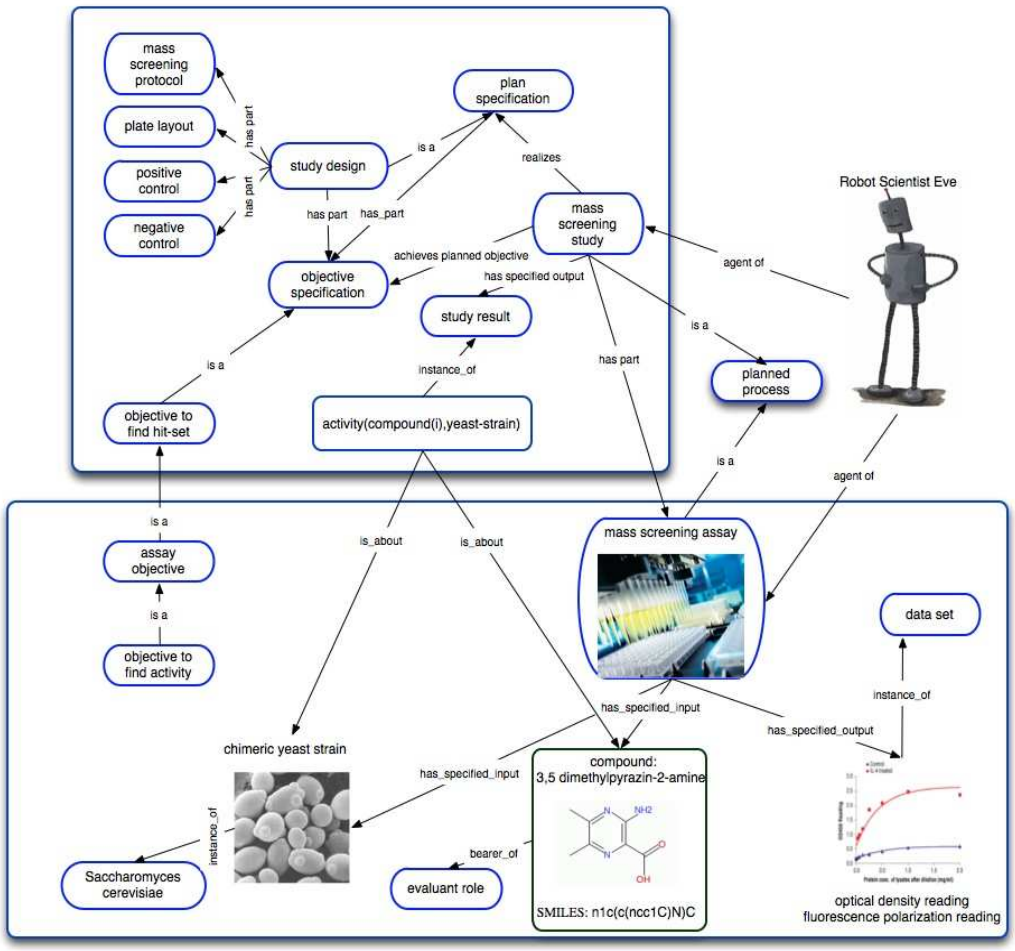

Figure 4: A fragment of a mass screening study run by Eve.

In this section we describe one blue box “mass screening study” from the Figure 3 on more detailed level. We employ the same methodology which we used for the description of investigations in LABORS [8] and for modelling the experimental processes in OBI.

During an investigation into development of drugs for a certain disease and for defined targets, Eve identifies what chimeric yeast strain to use for screening against a compound library. Eve initiates a *ddi:mass screening study* which is modelled as an *obi:planned process* (Figure 4). The study *realizes* an *obi:plan specification* which specifies a *ddi:objective to find hit-set* and an *obi:study design*. The study design consists of a *ddi:mass screening protocol*, specification of positive and negative control, and a *ddi:plate layout* that defines which wells contain yeast and compounds, and which wells are compounds free.

The mass screening study *has part* another planned process *ddi:mass screening assay* which *achieves planned objective* to identify activity in the compounds. The assay has such *obi:material entity* as compound which is a *bearer of obi:evaluant role* and yeast as specified in-

put, and it outputs optical density and fluorescence polarization reading which are modelled as *obi:data set*. These data are analysed by Eve and conclusions about hits are made which are modelled as *obi:study result*.

4 Conclusion and future work

DDI proposes a framework for unambiguous and formalised description of drug discovery investigations. DDI has been developed to support the Robot Scientist Eve which is designed to run automatic drug discovery investigations. In the development of the proposed ontology, we have followed OBO Foundry principles and therefore DDI is fully compliant with OBO formalisms and can be easily integrated with other existing ontology resources. DDI is designed in such a way that it can be extended to support the full pipeline of drug discovery.

In the next stage of the DDI project we will collaborate with a number of research groups (e.g. the developers of VO) in order to extend and to integrate DDI with external resources so that it can support cross disciplinary queries. We plan to submit DDI to the OBO Portal. The adoption of DDI will improve the retrieval of past drug discovery investigations, and promote secondary data reuse. DDI supports ontology-oriented databases which are more flexible than relational ones. The use of DDI will also improve data curation and maintenance. Use of DDI will promote semantic web applications that improve lab automation, such as automatically recording experimental data in e-Lab notebooks. In conclusion, use of DDI will add value to the data and methods used by pharmaceutical companies, and improve the efficiency of drug discovery process.

Acknowledgements

This work is supported by BBSRC grant BB/E018025/1, BB/FOF/PF/13/09 and RC UK. We would like to thank Dr Wilma Keighley, Life Science Consulting, UK and Dr James Ronald Hudson, Technology Transfer Manager at Aberystwyth University, Wales for consultancy on the DDI project. We thank Dr Andrew Sparkes at Aberystwyth University, Wales for his contribution to the database development.

References

[1] Gaia V. Paolini, Richard H. B. Shapland, Willem P. van Hoorn, Jonathan S. Mason, and Andrew L. Hopkins. Global mapping of pharmacological space. *Nature Biotechnology*, 24(7):805–815, July 2006.

[2] Jason D. Hughes, Julian Blagg, David A. Price, Simon Bailey, Gary A. DeCrescenzo, Rajesh V. Devraj, Edmund Ellsworth, Yvette M. Fobian, Michael E. Gibbs, Richard W. Gilles, Nigel Greene, Enoch Huang, Teresa Krieger-Burkee, Jens Loesel, Travis Wager, Larry Whiteley, and Yao Zhang. Physiochemical drug properties associated with in vivo toxicological outcomes. *Bioorganic & Medicinal Chemistry Letters*, 18(17):4872–4875, September 2008.

[3] Michael R. Barnes, Lee Harland, Steven M. Foord, Matthew D. Hall, Ian Dix, Scott Thomas, Bryn I. Williams-Jones, and Cory R. Brouwer. Lowering industry firewalls: pre-competitive informatics initiatives in drug discovery. *Nature Reviews Drug Discovery*, 8 (9):701–708, July 2009.

[4] Eric Bonabeau, Alpheus Bingham, and Aaron Schacht. Getting the Drugs We Need. *Harvard Business Review*, pages 48–50, January–February 2010.

[5] The Gene Ontology Consortium. Gene Ontology: tool for the unification of biology. *Nature Genet.*, 25:25–29, 2000.

[6] Susanna-Assunta Sansone, Daniel Schober, Helen Atherton, Oliver Fiehn, Helen Jenkins, Philippe Rocca-Serra, Denis Rubtsov, Irena Spasic, Larisa Soldatova, Chris Taylor, Andy Tseng, Mark Viant, and Ontology Working Group Members. Metabolomics standards initiative: ontology working group work in progress. *Metabolomics*, 3(3):249–256, September 2007.

[7] Christopher Lipinski and Andrew Hopkins. Navigating chemical space for biology and medicine. *Nature*, 432(7019):855–861, December 2004.

[8] Ross D. King, Jem Rowland, Stephen G. Oliver, Michael Young, Wayne Aubrey, Emma Byrne, Maria Liakata, Magdalena Markham, Pinar Pir, Larisa N. Soldatova, Andrew Sparkes, Kenneth E. Whelan, and Amanda Clare. The Automation of Science. *Science*, 324(5923):85–89, April 2009.

[9] Ross D. King, Kenneth E. Whelan, Ffion M. Jones, Philip G. K. Reiser, Christopher H. Bryant, Stephen H. Muggleton, Douglas B. Kell, and Stephen G. Oliver. Functional genomic hypothesis generation and experimentation by a robot scientist. *Nature*, 427(6971): 247–252, January 2004.

[10] Andrew Sparkes, Wayne Aubrey, Emma Byrne, Amanda Clare, Muhammed Khan, Maria Liakata, Magdalena Markham, Jem Rowland, Larisa Soldatova, Kenneth Whelan, Michael Young, and Ross King. Towards Robot Scientists for autonomous scientific discovery. *Automated Experimentation*, 2(1):11, 2010. ISSN 1759-4499.

[11] Steffen Schulze-Kremer. Ontologies for Molecular Biology. In *Proceedings of the 3rd Pacific Symposium on Biocomputing*, pages 693–704, 1998.

[12] Barry Smith, Michael Ashburner, Cornelius Rosse, Jonathan Bard, William Bug, Werner Ceusters, Louis J Goldberg, Karen Eilbeck, Amelia Ireland, Christopher J Mungall, The OBI Consortium, Neocles Leontis, Philippe Rocca-Serra, Alan Ruttenberg, Susanna-Assunta Sansone, Richard H Scheuermann, Nigam Shah, Patricia L Whetzel, and Suzanna Lewis. The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration. *Nature Biotechnology*, 25:1251–1255, November 2007.

[13] Barry Smith, Werner Ceusters, Bert Klagges, Jacob Köler, Anand Kumar, Jane Lomax, Chris Mungall, Fabian Neuhaus, Alan L Rector, and Cornelius Rosse. Relations in biomedical ontologies. *Genome Biology*, 6(R46), April 2005.

[14] Larisa N. Soldatova and Ross D. King. An ontology of scientific experiments. *J. R. Soc. Interface*, 3(11):795–803, 2006.

[15] Paula de Matos, Rafael Alcantara, Adriano Dekker, Marcus Ennis, Janna Hastings, Kenneth Haug, Inmaculada Spiteri, Steve Turner, and Christoph Steinbeck. Chemical Entities of Biological Interest: an update. *Nucl. Acids Res.*, page gkp886, 2009.

[16] Mélanie Courtot, Frank Gibson, Allyson L. Lister, James Malone, Daniel Schober, Ryan R. Brinkman, and Alan Ruttenberg. MIREOT: The Minimum Information to Reference an External Ontology Term. In *ICBO: International Conference on Biomedical Ontology*, 2009.