Drug Discovery Today • Volume 00, Number 00 • May 2016 PERSPECTIVE

ELSEVIER

feature

Is preclinical data sharing the new norm?

Q1 Katharine A. Briggs

Q2 **Is preclinical data sharing the new norm? In my experience, it is certainly becoming more commonplace. However, it is not yet standard practice and remains the preserve of special projects. Here, I expound the benefits of sharing proprietary preclinical data using examples of successful initiatives. The main barriers to data sharing are then described, with suggestions for how these might be overcome. To maximise the benefits and minimise the risks involved, I suggest that organisations look to develop standard operating procedures for data sharing.**

Introduction

Q3 In this article, I look at the benefits and risks associated with sharing preclinical data. The proprietary nature of these data means that initiatives to share these data tend to be restricted to the scientific community or to closed groups. In my experience, preclinical data sharing is becoming more commonplace in the pharmaceutical industry. However, it is not yet standard practice and remains the preserve of special projects. If organisations wish to maximise the benefits and minimise the costs involved in data sharing, then they need to integrate this process into business as usual. Clear roles and responsibilities as to who owns the data, who can authorise data sharing, on what basis, and the steps needed to obtain authorisation would help simplify the whole process. In practical terms, there is a need for standard operating procedures for data sharing within the industry. Standards are fundamental to data sharing at all levels, from administrative and management, software, infrastructure, to data and scientific standards [1].

E-mail address: Katharine.briggs@lhasalimited.org.

Reasons for data sharing

Benefits to public health

Approximately 35% of all drug development projects fail as a result of toxicity detected during preclinical safety studies [2]. Therefore, animal studies are an important safeguard for the safety of patients. The results of preclinical studies are sometimes published as a condensed summary for the few compounds that reach the market. However, most of these data are not routinely published or shared in public databases owing to the confidential nature of the research that generates the data. This lack of dissemination is clearly not in the best interest of public health. Sharing of these data would allow comparison of new structures to already existing data and help avoid costly failures in the future.

Required to by funding organisation or journal publisher

Potential benefits to public health and the demand for greater transparency are also driving funding organisations to encourage data sharing. The Public Health Research Data Forum brings together more than 20 funders of global health research and aims to increase the availability and use of health research data in ways that are equitable, ethical, and efficient (http://grants.nih.gov/grants/policy/data_sharing/; www.mrc.ac.uk/research/research-policy-ethics/data-sharing/policy/; www.wellcome.ac.uk/About-us/Policy/Policy-and-position-statements/WTX035043.htm) [3]. In addition, many journal publishers now encourage or even require supporting data to be deposited in publicly available repositories or made available as supplementary material to the paper [4].

Reduction in animal testing

A key benefit of data sharing is a potential reduction in animal testing, one of the 3Rs (reduction, refinement, and replacement). Data and knowledge gained could also enable more informed decisions about what substances to test and what tests to perform. A reduction in animal testing brings with it additional savings in terms of the time and costs involved.

An initiative led by the National Centre for the Replacement, Refinement and Reduction of Animals in Research (NC3Rs) and the Medicines

1359-6446/© 2016 Published by Elsevier Ltd.
http://dx.doi.org/10.1016/j.drudis.2016.05.003

Please cite this article in press as: Briggs, K.A. Is preclinical data sharing the new norm?, Drug Discov Today (2016), http://dx.doi.org/10.1016/j.drudis.2016.05.003

and Healthcare Products Regulatory Agency (MHRA) involving 32 organisations sharing data for 137 compounds and 259 studies identified that the use of recovery animals could be reduced by up to 66%, saving thousands of animals globally each year [5,6].

Lhasa Limited (www.lhasalimited.org/) is a not-for-profit charitable organisation that exists to promote the sharing of data and knowledge in chemistry and the life sciences. It is responsible for running several proprietary data-sharing initiatives [7]. One of these groups was set up by the Dog Project [8]; partners involved in the project share data from single and repeat-dose studies on the effects of excipients used in pharmaceutical formulations. Information on the absence or presence of biological effects of particular vehicles is important for adequate planning of studies in animals. Feedback from participants in the project suggests that the database has been useful in deciding which vehicles to use and in preventing undesirable formulations being investigated [9,10].

Required to by regulator

Regulators recognise that animal testing needs to be kept to a minimum while still protecting humans and the environment. The US Food and Drug Administration (FDA) has granted several organisations access to data held in their archives via cooperative research and development agreements (CRADAs) or research collaboration agreements (RCAs). Access to these data enabled Lhasa Limited to create several new structural alerts for nephrotoxicity and reprotoxicity in the Derek Nexus expert system for toxicity prediction, which is now shared with the wider Derek Nexus membership [11–13].

A similar arrangement was initiated in April 2003 with the National Institute of Health Sciences (www.nihs.go.jp/english/index.html) in Japan to share their data to improve Derek Nexus structural alerts for prediction of mutagenicity, *in vivo* chromosome damage, and *in vivo* liver repeat-dose toxicity [14].

There have also been several projects involving single organisations sharing data with Lhasa Limited. For instance, NV Organon and Bayer were concerned that the mutagenicity alert for steroids was too generic, classifying all compounds with a steroid backbone as mutagenic, whereas they had evidence of structural features that prevented this activity. They chose to share this information with Lhasa Limited to improve predictive performance so that regulators would not insist that they tested every steroid and were happy to be named as the source of the data. In fact, around 25% of the structural alerts included in Derek Nexus have been developed using donations of confidential data.

Ethos of science

A case for data sharing can also be made on the basis of the ethos of science described by Robert Merton [15], whereby scientific findings should be made available to the entire scientific community to allow other researchers to conduct their own analyses and verify the results. Independent replication of research findings is seen as the fundamental mechanism by which scientific evidence accumulates to support a hypothesis.

In direct conflict with this is the need within industry to protect proprietary information. However, organisations need to be clear about how much of a competitive advantage they will lose by sharing data versus the knowledge they will additionally gain. How unique is the knowledge they hold versus the knowledge their competitors could bring to the table? Consideration should also be given to the risk of not taking part in data sharing because those organisations that participate will have a competitive and economic advantage. For example, if there are ten members in the group, then donating just a single test will give a return of ten test results.

Access to a bigger pool of data

Often legacy data is 'locked' inside PDFs sitting in individual company archives, where it is unavailable even for internal analysis. Providing access to a larger pool of data can reveal patterns that are simply not visible in smaller component data sets, where such relations might be represented by only one or two chemicals. Often, only regulatory bodies have ready access to pooled data sets and, therefore, have the opportunity to identify these broader patterns by performing cross-company analyses. This can present problems when submitting a new drug application because broader regulatory knowledge can lead to challenges and assertions that need to be addressed, resulting in delays and the need for additional data generation.

It could be argued that pooling data also presents a risk of uncovering undetected safety signals. However, it is unlikely these would remain undiscovered in the long term and learning about these later during the drug development pipeline will be more costly. Sharing data could also point to ways in which the drug candidate could be modified to remove safety issues.

One of the other data-sharing groups that Lhasa Limited are managing was set up in response to registration requirements for pharmaceuticals on the assessment of toxicological hazards presented by impurities stemming from the manufacturing process [16]. Often, this hazard assessment is done through the use of structural alerts and read-across determinations from available data sources. Therefore, expanding the available data pool offers significant benefits. To date, the partners in this collaboration have shared data on the mutagenicity of 969 noncommercially sensitive structures, such as starting material and intermediates.

External experts can offer fresh insight

Research data can be valuable many years after they have been generated and fresh eyes can reveal new insights beyond those originally identified. In addition, new research topics and fields are emerging between the boundaries of traditional disciplines. By sharing data, companies can gain from external expertise in the same or different fields, opening up the data to be explored and used in ways that might not have originally been envisioned. Given that the costs of generating the data are also shared, it opens the possibility for exploratory research that otherwise might not be commercially viable.

The Mechanism-Based Integrated Systems for the Prediction of Drug-Induced Liver Injury project (www.mip-dili.eu/index.php?page=home) aims to identify and validate an improved panel of *in vitro* best-practice assays for predicting drug-induced liver injury (DILI) in humans [17]. To achieve this objective, a database of public and proprietary information is being collated that will allow the consortium to make robust and objective decisions on training and test compounds to assess the *in vitro* models being developed.

The aryl boronic acids data-sharing group was initiated as a result of a presentation at the UK Environmental Mutagen Society (UKEMS) meeting [18]. This highlighted a knowledge gap relevant to most pharmaceutical companies that use aryl boronic acids mainly as reagents for C–C bond formation in drug development. The resulting consortium of seven participants has shared data for 18 commonly used aryl boronic acids tested in Ames mutagenicity, mouse lymphoma, *in vitro* micronucleus, and comet assays. This led to the development of a specific structural alert that was released to Derek Nexus members in 2012, although exclusion criteria are limited because of the unknown mechanism of mutagenicity [19].

A similar consortium for the investigation of genotoxicity of aromatic amines (CIGAA) arose as the result of a series of precompetitive knowl-

Please cite this article in press as: Briggs, K.A. Is preclinical data sharing the new norm?, Drug Discov Today (2016), http://dx.doi.org/10.1016/j.drudis.2016.05.003