1
Introduction to Sample Management

William P. Janzen and Andy Zaayenga

At its simplest level sample management is just inventory – where can one find a given item and retrieve it? But in the context of modern discovery efforts, be they drug discovery, agricultural, protein therapeutic, biobanking, or the plethora of other disciplines that collect and manage samples, the problem is far more complex. Today, sample management may have to manage millions of samples in a library that spans several continents but will also have to contend with a worldwide customer base. To make the problem more difficult the content of the samples must also be managed, which may involve complex chemical structures and storage conditions that may vary from room temperature under inert atmosphere to storage in liquid nitrogen. At this level the storage of these samples must now involve complex informatics and automation systems. This volume will capture the best practices compiled from experts in the field of sample management and will hopefully serve as a guide to both novice sample managers who need to track a few thousand compounds in room-temperature vials to professionals in multinational organizations.

As long as there have been chemicals there has been a need for sample management. One could imagine that for a seventeenth century druggist this was simply an inventory of the herbal extracts and remedies he compounded into salves and potions and the location where they were stored. This could be done from memory in most cases and probably evolved to a written inventory when searching for needed components became too slow and cumbersome. Early sample management evolved in parallel with drug discovery. What we consider sample management today came into being as pharmaceutical companies began to amass chemical libraries and test these in disease-focused assays. As these companies synthesized compounds, they retained samples and began to amass collections of chemical compounds that numbered in the tens of thousands. At the same time, the testing of natural product extracts became common practice, significantly boosting the number of samples to be stored [1, 2]. As the number of samples exceeded 100 000 (at that time a seemingly immense number), automated systems were developed to store and catalog them. Initially, these were simple robotic units or adapted card file systems that would simply present entire drawers or boxes of samples to an operator. Chemical structures were often still paper copies and stored

Management of Chemical and Biological Samples for Screening Applications, First Edition.
Edited by Mark Wigglesworth and Terry Wood.

elsewhere, and the amount in the inventory was rarely accurate if tracked at all. Storage labware formats were standardized to accommodate the large volumes of samples moving through the system and to facilitate liquid handling and detection platform development [3]. Improvements in liquid handling and detection enabled increasingly higher labware densities, allowing tighter environmental control, and larger libraries.

Sample integrity became paramount with a focus on environmental conditions and consistent sample history both in the storage units and in the reformatting/analysis areas. Significant numbers of legacy compounds which had been subjected to variable temperatures, water, oxygen, and light were found to be compromised. Container seal adhesives and labware mold components could introduce interferents to the assay results. Compound managers realized that consistent sample quality was a key to valid scientific data. Programs were employed to provide cradle-to-grave care as well as purity monitoring to insure repeatable sample integrity.

With the advent of combinatorial chemistry, parallel synthesis made the creation of large compound sets numbering in the hundreds of thousands viable and raised the stakes for compound management. High-throughput screening (HTS) groups began requiring that compounds be presented in 96 well plates dissolved in Dimethyl sulfoxide (DMSO) and consumed these plates at an alarming rate. At about the same time, the electronic storage and representation of chemical structures became possible [4]. As the numbers of samples increased, chemists could no longer rely on visual inspection of structures, so tools were developed to analyze synthetic sets to determine their degree of similarity or difference [5]. Compound management groups, that had often become underfunded corporate backwaters, suddenly found themselves under the spotlight as the bottleneck in an exciting new process.

In answer to this challenge, funding was allocated to revamp chemical stores, and a plethora of bespoke systems of automation and data management appeared [1, 6–8]. The linkage between automated preparation systems and data systems was a slow process and the systems that were created varied widely in their architecture and success but shared a number of traits that embody today's samples management system:

- Sample registration
- Usage of enterprise-wide standardized labware
- Positive sample tracking, usually using barcodes
- Cradle-to-grave tracking of samples stored in both vials and plates
- Sample security with user access tracking and control
- Accurate quantity tracking of both mass and volume
- Storage of compounds in DMSO solutions
- Control of environmental conditions to minimize water uptake, oxygen and light degradation, and temperature fluctuation
- High speed automated storage and retrieval
- Reduced freeze/thaw cycles by efficient daughter plate production or by multiple aliquotting

- Regular purity monitoring
- Sample ordering systems
- Automated cherry picking and plate preparation systems
- Robust distribution methods and sample tracking outside of the storage system.

As HTS and ultra high throughput screening (uHTS) became ubiquitous tools in drug discovery, they also began to be used in other industries such as the discovery of agricultural agents (pesticides, animal health, etc.), catalysts, polymer discovery, fragrances, and flavorings. Similarly, the newly created science of sample management also found utility in many other areas. With the advent of the human genome project the need to store large numbers of biological samples became imperative [9]. Blood and tissue sample banks both for research and distribution grew to the point where they required similar techniques (Table 1.1). Today sample management is applied in industries as varied as hospital pharmacies, environmental repositories, and sperm banks. So let us now examine the techniques used in modern sample management.

Inventory: Probably the most critical factor in sample management remains inventory. But this has expanded well beyond the simple 'where is it' definition. Today's sample manager is more concerned with curation of the samples in their charge. As is discussed in Chapters 2–4, this includes the integrity of the samples on receipt, during storage, and even after delivery to end customers. To accomplish this, samples must be subjected to analytical tests for quality control (QC). In many cases this information will be provided by the supplier of the material. When that supplier is an internal group or a trusted partner this may be deemed

Table 1.1 Comparison of compound management and biobanking.

Compound management	Biobanking
Compounds are precious but for the most part replaceable	Specimens irreplaceable
Freeze/thaw cycles kept low, target 6–10	Freeze/thaw cycles very low or none
Large legacy libraries to be automated, which were added to en masse through library purchase or acquisitions/mergers	Small collections, few legacy specimens to be automated, samples added incrementally
Unregulated environment, compliance requirements low	Regulated environment, compliance requirements high
Low probability of cross organization exchange	High probability of cross organization exchange
Historically large budgets for R&D, low examination of return on investment (ROI), long-term funding available	Costs and resources may be subsidized, budgets and ROI examined closely, long term financing to cover length of studies difficult
Quality of legacy samples questionable	Quality of legacy samples questionable

sufficient and accepted but in other cases the purity of the material will have to be verified. For small molecule samples the most common method applied is Liquid Chromatography/Mass Spectrometry (LC/MS) analysis [10, 11].

In biobanking, the focus is on maintaining quality from collection to analysis. Here the primary problem is that one cannot sample the specimen regularly due to degradation during the aliquoting process. Also, the specimen volume is likely to be very small and therefore prized. Establishing the purity on receipt is a critical first step but is rarely sufficient. The quality of the samples in storage must be verified over time and, in many cases, after dissolution. The latter remains a largely unsolved problem as of the time of this writing. Dissolving a compound introduces a host of QC problems, particularly when the samples are transferred to plates. While it is possible to test the concentration and purity of samples dissolved in DMSO, it is not practical to test hundreds of thousands of samples on a regular basis using these techniques. In addition, it is nearly impossible to test small-molecule chemical samples in the environment used for HTS. Representative sampling of libraries has shown that a relatively high proportion (>20%) of the compounds in a sample set will be insoluble after a simple water dilution [12]. On the other hand, empirical data shows that many of these compounds will show activity in certain buffer or cellular testing systems implying that they are soluble under alternative conditions. The solution that many laboratories have adopted is to test subsets of the library and to test compounds that are determined to be active and are requested for further follow up. This approach is discussed in more detail in Chapters 2 and 15.

To make the problem even more complex, the samples may be subjected to various storage conditions and may be shipped to alternative sample management sites or end customers. The number of times the sample has been frozen and thawed and the storage temperatures may affect the stability of the sample set. There is not a clear body of literature on sample stability in DMSO [13, 14] and conflicting anecdotal evidence making the choice of storage conditions for DMSO samples difficult. The unusual physical properties of DMSO also complicate this matter [15]; DMSO will readily absorb water and oxygen from the atmosphere, which radically changes its freezing point and may affect the stability of compounds. The range of approaches in this area is widely varied. Some groups have established maximum freeze/thaw ranges and employ single-use plates in their process to minimize atmospheric exposure, while others have embraced room temperature storage and accepted the inevitability of water uptake by adding 10–20% water to their DMSO prior to sample dissolution [16–18]. This broad range of approaches and the fact that all have produced lead compounds makes establishing a true best practice impossible.

Tracking the location and history of samples is neither simple nor taken for granted. The use of barcodes is ubiquitous in sample management today. Barcodes are, in essence, very simple; they are simply a way of recording a serial number and rapidly and accurately entering that into a computer system. Barcode-based inventory systems, on the other hand, can be quite complex [19, 20]. They require the assignment of a tracking ID to every sample and a complex data model to register every manipulation of a sample from weighing through solubilization and

any transfer from container to container. This system must always have not only the current volume of every sample but the historical record of every transfer from the lot submitted to the disposal of the last plate.

Data systems supporting sample management are discussed in Chapters 13 and 14. In addition to inventory, they will usually incorporate some mechanism for managing requests for samples. Customer ordering systems should always appear simple to end customers but may have quite complex internal management structures for the sample management professional. This can include 'pull' systems, where customer ordering software allows users to request samples and specify form (i.e., solid or liquid and concentration) and even location on a plate. Other aspects of a system may employ 'push' systems that automatically assign work to be performed on samples. For example, chemical samples that are synthesized as part of a specific medicinal chemistry program will usually have a prescribed group of assays that need to be conducted on each compound. When a chemist submits compounds, he or she may associate the molecules with a given program, and the IT system will automatically create work orders that create plates, tubes, and/or vials that are routed to the laboratories performing these tests.

The final aspect of sample management is automation systems, found in Chapters 10 and 12. While the management of samples does not require automation, it is virtually impossible to support the management of a large library without some degree of automation. Automated systems can range from simple liquid handling units that perform vial-to-plate transfers and plate-to-plate replication to large fully integrated systems that can perform all the aspects of sample preparation from sample dissolution to final microplate preparation. It should be noted that the one aspect of sample management that has never been efficiently automated is the weighing of samples. While significant resources have been devoted to this problem, automated solutions have been stymied by the highly varied nature of the chemical samples themselves. These samples can range from very dry, free flowing powders (which are easy to dispense) to tars that must be scraped or transferred by dipping a spatula, or proteins that are extremely hygroscopic and form light flakes that blow away easily. As a result almost all laboratories still employ a manual weighing process that is highly integrated with a data system and sample tracking to ensure accuracy – this is another reason that QC is so important. As with many HTS applications, this aspect of sample management has largely been solved. The systems have evolved from gymnasium-sized units that could store 1 million samples and process 10 000–20 000 samples per day to small unit stores that can be connected to provide the same storage capacity in a standard laboratory. Similarly, the problem of low-temperature storage has largely been solved. Systems that operate at temperatures down to vapor phase liquid nitrogen storage are now available and are discussed in the biobanking sections of this volume. Modern sample management systems enable the automated storage of virtually any sample from small-molecule chemicals to cells and tissues.

So, in conclusion, the field of sample management has grown both in importance and sophistication. The importance of this activity cannot be underestimated. The cost to replace a corporate chemical collection can be conservatively estimated at

$500 per sample. For a 1 million compound collection this might take four to five years and cost $500 000 000. Additionally, human and non-human biological specimens are irreplaceable. Even if a replacement specimen can be obtained, the biological state will have changed and the specimen will not be identical. When looked at in this light, it would be criminal to allow compounds or specimens to degrade or be lost.

References

1. Archer, J.R. (2004) History, evolution, and trends in compound management for high throughput screening. *Assay Drug Dev. Technol.*, **2** (6), 675–681.
2. Janzen, W.P. and Popa-Burke, I.G. (2009) Advances in improving the quality and flexibility of compound management. *J. Biomol. Screen.*, **14** (5), 444–451.
3. ANSI (2004) *New Microplate Standards Expected to Accelerate and Streamline Industry*, ANSI, New York, *http://www.ansi.org/news_publications/news_story.aspx?menuid=7&articleid=598* (accessed 2004). ANSI/SBS 1-2004: Footprint Dimensions ANSI/SBS, 2-2004: Height Dimensions ANSI/SBS, 3-2004: Bottom Outside Flange Dimensions ANSI/SBS 4-2004: Well.
4. Warr, W.A. (1991) Some observations on piecemeal electronic publishing solutions in the pharmaceutical industry. *J. Chem. Inf. Comput. Sci.*, **31** (2), 181–186.
5. Oprea, T.I. (2000) Property distribution of drug-related chemical databases. *J. Comput. Aided Mol. Des.*, **14** (3), 251–264.
6. Rutherford, M.L. and Stinger, T. (2001) Recent trends in laboratory automation in the pharmaceutical industry. *Curr. Opin. Drug Discov. Devel.*, **4** (3), 343–346.
7. Ray, B.J. (2001) Value your compound management team! *Drug Discov. Today.*, **6** (11), 563.
8. Janzen, W.P. (2002) *High Throughput Screening: Methods and Protocols*, Humana Press, Totowa, NJ.
9. Eiseman, E. and Haga, S. (1999) *Handbook of Human Tissue Sources: A National Resource of Human Tissue Samples*, Rand Corporation.
10. Ari, N., Westling, L., and Isbell, J. (2006) Cherry-picking in an orchard: unattended LC/MS analysis from an autosampler with >32,000 samples online. *J. Biomol. Screen.*, **11** (3), 318–322.
11. Letot, E., Koch, G., Falchetto, R., Bovermann, G., Oberer, L., and Roth, H.J. (2005) Quality control in combinatorial chemistry: determinations of amounts and comparison of the 'purity' of LC-MS-purified samples by NMR, LC-UV and CLND. *J. Comb. Chem.*, **7** (3), 364–371.
12. Popa-Burke, I.G., Issakova, O., Arroway, J.D., Bernasconi, P., Chen, M., Coudurier, L. *et al.* (2004) Streamlined system for purifying and quantifying a diverse library of compounds and the effect of compound concentration measurements on the accurate interpretation of biological assay results. *Anal. Chem.*, **76** (24), 7278–7287.
13. Kozikowski, B.A., Burt, T.M., Tirey, D.A., Williams, L.E., Kuzmak, B.R., Stanton, D.T. *et al.* (2003) The effect of freeze/thaw cycles on the stability of compounds in DMSO. *J. Biomol. Screen.*, **8** (2), 210–215.
14. Kozikowski, B.A., Burt, T.M., Tirey, D.A., Williams, L.E., Kuzmak, B.R., Stanton, D.T. *et al.* (2003) The effect of room-temperature storage on the stability of compounds in DMSO. *J. Biomol. Screen.*, **8** (2), 205–209.
15. Rasmussen, D.H. and Mackenzie, A.P. (1968) Phase diagram for the system water-dimethylsulphoxide. *Nature*, **220** (5174), 1315–1317.
16. Schopfer, U., Engeloch, C., Stanek, J., Girod, M., Schuffenhauer, A., Jacoby, E. *et al.* (2005) The Novartis compound

archive – from concept to reality. *Comb. Chem. High Throughput Screen.*, **8** (6), 513–519.

17. Jacoby, E., Schuffenhauer, A., Popov, M., Azzaoui, K., Havill, B., Schopfer, U. *et al.* (2005) Key aspects of the Novartis compound collection enhancement project for the compilation of a comprehensive chemogenomics drug discovery screening collection. *Curr. Top. Med. Chem.*, **5** (4), 397–411.
18. Engeloch, C., Schopfer, U., Muckenschnabel, I., Le Goff, F., Mees, H., Boesch, K. *et al.* (2008) Stability of screening compounds in wet DMSO. *J. Biomol. Screen.*, **13** (10), 999–1006.
19. Palmer, R.C. (1995) *The Bar Code Book: Reading, Printing, Specification, and Application of Bar Code and Other Machine Readable Symbols*, 3rd edn, Helmers Publishing Inc., Peterborough, NH.
20. Burke, H.E. (1990) *Automating Management Information Systems*, Van Nostrand Reinhold, New York.