



# EPA Public Access

Author manuscript

*Regul Toxicol Pharmacol.* Author manuscript; available in PMC 2019 October 28.

About author manuscripts

Submit a manuscript

Published in final edited form as:

*Regul Toxicol Pharmacol.* 2017 December ; 91(Suppl 1): S27–S35. doi:10.1016/j.yrtph.2017.10.007.

## Framework for the quality assurance of 'omics technologies considering GLP requirements

Hans-Martin Kauffmann<sup>1</sup>, Hennicke Kamp<sup>1</sup>, Regine Fuchs<sup>2</sup>, Brian N. Chorley<sup>3</sup>, Lize Deferme<sup>4</sup>, Timothy Ebbels<sup>5</sup>, Jörg Hackermüller<sup>6</sup>, Stefania Perdichizzi<sup>7</sup>, Alan Poole<sup>8</sup>, Ursula G. Sauer<sup>9</sup>, Knut E. Tollefsen<sup>10</sup>, Tewes Tralau<sup>11</sup>, Carole Yauk<sup>12</sup>, Ben van Ravenzwaay<sup>1,\*</sup>

<sup>1</sup>BASF SE, Germany

<sup>2</sup>metanomics GmbH, Germany

<sup>3</sup>U.S. Environmental Protection Agency, USA

<sup>4</sup>ExxonMobil Petroleum and Chemical B.V.B.A., Belgium

<sup>5</sup>Imperial College London, United Kingdom

<sup>6</sup>Department of Molecular Systems Biology, Helmholtz Centre for Environmental Research - UFZ, Germany

<sup>7</sup>Center for Environmental Toxicology, Agency for Prevention, Environment and Energy (Arpae), Emilia-Romagna, Italy

<sup>8</sup>European Centre for Ecotoxicology and Toxicology of Chemicals (ECETOC), Belgium

<sup>9</sup>Scientific Consultancy – Animal Welfare, Germany

<sup>10</sup>Norwegian Institute for Water Research (NIVA), Norway

<sup>11</sup>Department of Chemical and Product Safety, German Federal Institute of Risk Assessment (BfR), Germany

<sup>12</sup>Environmental Health Science and Research Bureau, Health Canada, Canada

### Abstract

'Omics technologies are gaining importance to support regulatory toxicity studies. Prerequisites for performing 'omics studies considering GLP principles were discussed at the European Centre for Ecotoxicology and Toxicology of Chemicals (ECETOC) Workshop *Applying 'omics technologies in Chemical Risk Assessment*. A GLP environment comprises a standard operating

\* **Corresponding author:** Ben van Ravenzwaay, BASF SE, RB/T - Z470, 67056 Ludwigshafen, Germany, Phone: +49 621 60-56419; Fax: +49 621 60-58134, bennard.ravenzwaay@basf.com.

<sup>6</sup>Conflict of interest

UGS was hired by ECETOC and the European Chemical Industry Council Long-range Research Initiative (CEFIC LRI) to assist in the preparation of the manuscript. The other authors were engaged in the course of their normal employment. The authors alone are responsible for the content and writing of the paper.

<sup>7</sup>Disclaimer

The content described in this article has been reviewed by the National Health and Environmental Research Laboratory of the U.S. Environmental Protection Agency and approved for publication. Approval does not signify that the contents necessarily reflect the views and the policies of the Agency. Mention of trade names or commercial products does not constitute endorsement or recommendation for use.

procedure system, proper pre-planning and documentation, and inspections of independent quality assurance staff. To prevent uncontrolled data changes, the raw data obtained in the respective 'omics data recording systems have to be specifically defined. Further requirements include transparent and reproducible data processing steps, and safe data storage and archiving procedures. The software for data recording and processing should be validated, and data changes should be traceable or disabled. GLP-compliant quality assurance of 'omics technologies appears feasible for many GLP requirements. However, challenges include (i) defining, storing, and archiving the raw data; (ii) transparent descriptions of data processing steps; (iii) software validation; and (iv) ensuring complete reproducibility of final results with respect to raw data. Nevertheless, 'omics studies can be supported by quality measures (e.g., GLP principles) to ensure quality control, reproducibility and traceability of experiments. This enables regulators to use 'omics data in a fit-for-purpose context, which enhances their applicability for risk assessment.

## Keywords

Good laboratory practice (GLP); standard operating procedure; independent quality assurance; documentation; data storage; reproducibility; raw data definition; software validation; quality assurance inspection

## 1 Introduction

'Omics technologies, such as genomics, proteomics, metabolomics, and transcriptomics, are rapidly developing research technologies, and they are gaining increasing importance to support regulatory toxicity studies. The application and integration of 'omics technologies may be useful in different layers of regulatory hazard identification and assessment contributing to (i) the classification and labelling of substances, for example as part of a tiered testing strategy; (ii) weight-of-evidence approaches to elucidate the modes-of-action of the substance under investigation; (iii) the substantiation of chemical similarity for read-across (ECHA, 2015; van Ravenzwaay et al., 2016); (iv) the determination of points-of-departure for hazard assessment; (v) the demonstration of species-specific effects and human health relevance (or absence thereof). Therefore, studies including 'omics technologies could make an important contribution to the risk assessment of substances (Buesen et al., 2017; Sauer et al., 2017).

Generally, regulatory toxicity studies must be performed according to principles of good laboratory practice (GLP) if they are intended to fulfil legal requirements to support the notification or regulatory approval of substances. Exceptions are, e.g., some specialized assays within the scope of immunotoxicity testing of pharmaceuticals, for which it is accepted that they might not comply fully with GLP (EMA, 2006). The Organisation for Economic Cooperation and Development (OECD) has provided a general GLP framework in its *Principles of good laboratory practice and compliance monitoring* (OECD, 1998) and a number of related OECD GLP consensus, guidance and advisory documents.

In order to use the data obtained in 'omics-based studies for regulatory purposes, it would be beneficial to also conduct these investigations according to the principles of GLP. This would serve the goals (i) to promote the consistent quality and validity of data used for

determining the safety of chemical products, a primary objective of the GLP principles (OECD, 1998); (ii) to promote transparent process descriptions and thus support the traceability of study results; and (iii) to facilitate the exchange of information and enhance the regulatory impact of ‘omics data, if successfully used for hazard and risk assessment purposes. All of these issues are expected to enhance the applicability of ‘omics data in a regulatory context, but also in research consortia where different project partners use the same data for different purposes.

Moreover, the *OECD Council Decision on the mutual acceptance of data* (OECD, 1981) states that (eco)toxicological test data generated in any OECD member country in accordance with OECD Test Guidelines and the principles of GLP shall be accepted in other member countries. Hence, adherence to the principles of GLP also facilitates the mutual acceptance of data, and the GLP status of data avoids duplicate testing for different authorities.

‘Omics technologies are part of a fast-growing scientific field that has a primary focus on the investigation of research questions. In this area, the GLP status of data is not required. However, as ‘omics technologies are refined and knowledge on ‘omics increases, the incentive to address questions specifically related to hazard identification and assessment using ‘omics technologies is increasingly gaining importance. It is in this regulatory context that GLP conditions are relevant. However, to date, guidance is unavailable on how to conduct ‘omics studies considering the principles of GLP. Importantly, for some steps of ‘omics studies, full compliance with the principles of GLP would be very difficult or even impossible to achieve due to specific technical features inherent to these steps. For ‘omics studies that encompass such steps, it appears to be a more realistic goal to strive for ‘GLP-like’ conditions.

‘GLP-like’ studies are defined as studies that are conducted in compliance with the principles of GLP as far as technically feasible, in which those steps that cannot be conducted in compliance with the principles of GLP for technical reasons are clearly identified and the specific technical reasons that stand in the way to full GLP compliance are clearly described.

Against this background, the establishment of a GLP(-like) context for collecting, storing and curating ‘omics data was one of the key objectives of the European Centre for Ecotoxicology and Toxicology of Chemicals (ECETOC) workshop *Applying ‘omics technologies in chemical risk assessment*, that took place on 10-12 October 2016 in Madrid, Spain (Figure 1). The report of this workshop is provided in Buesen et al. (2017) in this journal Supplement. Ahead of the ECETOC workshop, a first draft of a *Framework establishing a GLP(-like) context for collecting, storing and curating ‘omics data* was compiled, and this draft was presented and further discussed during work stream 1 of the workshop.

Data processing and data interpretation were not part of the GLP discussions, but were specifically addressed in work streams 2, 3, and 4 (Bridges et al., 2017; Gant et al., 2017; in this journal Supplement) (Figure 1). The participants of the workshop, who brought in

expertise with respect to transcriptomics / toxicogenomics, metabolomics and GLP, provided recommendations to advance the draft framework. These recommendations were addressed in updating the first draft, yielding the present article *Framework for the quality assurance of 'omics technologies considering GLP requirements*.

Generally, 'omics-based studies encompass three steps, i.e. (i) taking a tissue, blood, or cell sample from an *in vivo* or *in vitro* study; (ii) analysis using 'omics technologies; and (iii) scientific interpretation of the 'omics data (Figure 2). The concept to introduce GLP requirements for 'omics investigations raises certain challenges which are described and discussed below with respect to the different aspects that are relevant to conduct an 'omics study under GLP conditions (Section 2). Section 3 exemplarily presents the workflow of a metabolomics study to describe how GLP processes can be considered in an 'omics study. Section 4 summarizes the recommendations from the ECETOC workshop related to the quality assurance of 'omics technologies considering GLP requirements, and Section 5 provides a discussion of key issues addressed in this article and draws conclusions therefrom.

## 2 Application of the OECD principles of GLP to 'omics-based studies

Originally, the GLP principles were developed to promote the quality and validity of preclinical safety data. Test facilities performing GLP studies are regularly inspected for GLP compliance by national agencies based on national or agencies' GLP regulations. For some aspects (e.g. duration of archiving), these regulations can vary with agencies or nations, but the basic principles are very similar to ensure data quality and integrity. The GLP requirements established in national or agencies' regulations rely mainly on the *OECD Principles of good laboratory practice and compliance monitoring* (OECD, 1998) that serve to ensure that the evaluation of potential hazards of substances are based on safety data of sufficient quality, accuracy and reproducibility.

The OECD principles of GLP (OECD, 1998) cover specific chapters on (i) test facility organisation and personnel; (ii) quality assurance programme; (iii) facilities; (iv) apparatus, material, and reagents; (v) test systems; (vi) test and reference items; (vii) standard operating procedures (SOPs); (viii) performance of the study; (ix) reporting of study results; and (x) storage and retention of records and materials.

Of note, in 2004, the OECD published the *Advisory Document No. 14 on the application of the principles of GLP to in vitro studies* (OECD, 2004). This advisory document was developed in view of *anticipated developments in the fields of toxicogenomics, toxicoproteomics, toxicometabonomics and in various high throughput screening techniques* that were expected to enhance the importance of *in vitro* methodologies for safety testing (OECD, 2004).

As presented in further detail in the following Sub-sections, basic GLP principles that are relevant for the quality assurance of 'omics technologies considering GLP requirements, include:

- 2.1: Organisational aspects

- 2.2: Standard Operating Procedures (SOPs)
- 2.3: Study planning;
- 2.4: Definition of raw data
- 2.5: Data processing and storage
- 2.6 Reporting;
- 2.7: Software validation;
- 2.8: Archiving of data.

## 2.1 Organisational aspects

Establishing a GLP process for 'omics-based studies should be achievable if the toxicity studies used to source the tissues and/or blood samples for the 'omics analysis are already carried out in compliance with GLP principles and in the same GLP-certified laboratory or test facility. If the 'omics investigations are performed in a separate facility, it will be necessary to ensure that the required organisational environment for GLP studies is implemented there (*cf.* Section 2.3 for requirements related to multi-site studies). In addition to organisational requirements addressing basic principles and functions (e.g. test facility management, study director, archivist, and quality assurance staff), this comprises the establishment of a quality management programme as described in the OECD principles of GLP. Further, all people working in a GLP environment should be regularly trained in the principles of GLP, and they should be promptly informed of any relevant changes in the GLP requirements or quality management system.

## 2.2 Standard Operating Procedures (SOPs)

All basic processes related to GLP studies must be described in SOPs, which must be approved by the test facility management. If a facility starts working under GLP, such a SOP system will have to be established. The GLP principles and additional related international and national legislation or related provisions (if established) describe the specific areas and topics which at least have to be covered by SOPs.

Briefly, SOPs should describe the entire general process from study planning via experimental conduct and data recording up to report generation and archiving. Besides, they should comprise appropriate measures to ensure the correctness and reproducibility of results. All relevant laboratory preconditions should also be described in SOPs (e.g. use, maintenance, cleaning, calibration, validation of used apparatus or systems, acceptance criteria or tolerance ranges).

Hence, the corresponding general activities and preconditions for 'omics studies should be predefined and described in SOPs. This includes the general procedure for the analysis and processing of 'omics data. Study-specific details, on the other hand, have to be described in a study plan (*cf.* Section 2.3). Procedures described in SOPs and study plan/amendments should have the necessary level of detail and precision to ensure that the study investigator can conduct the study on the basis of these planning documents (if necessary together with

separate technical manuals for equipment or kits), because study plans and amendments as well as SOPs are the only planning documents stipulated in the GLP principles.

Whereas the recording, processing and archiving of data must follow predefined or standardised procedures, data interpretation and conclusions depend on the expertise of the investigator and the corresponding process can usually not be described in advance. Consequently, there is no GLP requirement to predefine the scientific assessment of generated data in SOPs.

### 2.3 Study planning

Whereas SOPs describe general procedures, the study plan describes study-specific details, such as study code and title, responsible staff, test item, test item concentrations, sample number and origin, and experimental design (e.g. study-specific methods and/or statistical analysis that shall be applied). The minimum contents of a study plan are also defined in the GLP principles.

The planning of separate 'omics studies or the integration of 'omics investigations in regulatory toxicity studies is considered feasible, given that the technology can be standardised in a manner that the experimental study design is fixed and can be properly described in a study plan (and/ or SOPs).

However, for complex study designs (e.g. genome-wide gene expression analysis) such a complete pre-planning may not always be possible because, e.g., results generated during the study can trigger specific further data processing or analysis procedures. In such cases, subsequent planning steps, which cannot be determined prior to the onset of the study, should be described and justified in consecutive study plan amendments to be fully GLP-compliant. Nevertheless, such a consecutive planning description could be quite laborious for huge and complex 'omics investigations which are difficult or even impossible to standardize.

For study planning, it should also be considered that the study director has to ensure that computerised systems used in his or her GLP study have been validated (*cf.* Section 2.7).

If the 'omics analysis is conducted under GLP conditions in a facility that is not identical with the facility performing the toxicity study, the study design can be set up as so-called GLP multi-site study (OECD, 2002). In multi-site studies, a single study director, who is usually located in the laboratory conducting the animal toxicity study, has the overall responsibility for conducting and evaluating the entire study, and this study director is supported by a principal investigator (PI) at the external test site conducting the 'omics analyses. This scenario requires excellent communication between the study director and the PI regarding the 'omics investigations, to ensure that the study plan comprises all relevant aspects of the (potentially very complex) 'omics part of the study and that the study director is promptly informed about deviations from the study plan or the urgency to issue study plan amendments.

## 2.4 Definition of raw data

GLP raw data are defined as *all original test facility records and documentation, or verified copies thereof, which are the result of the original observations and activities in a study* (OECD, 1998). For each data recording system used, the corresponding raw data must be defined. Hence, for each specific ‘omics technology, the actual raw data will most likely be different.

At the ECETOC workshop, there was general consensus that raw data obtained from ‘omics technologies would be best defined as the first set of data, obtained from the equipment without direct interference of the operator, that provides interpretable results (Buesen et al., 2017). Accordingly, the specific raw data depend on the technology and equipment applied:

- For transcriptomics data generated by gene expression arrays, raw data can constitute the files containing signal intensity readouts (also referred to as feature extraction files; e.g., CEL files for Affymetrix® platforms; and TXT or SHP files for Agilent platforms) that have not yet been submitted to the data normalisation process. At the workshop, it was further discussed that image data might also be considered as first raw data.
- For RNA-sequencing data, the raw data file for Illumina® sequencing is the Base Call File (\*.bcl file), which is the primary sequencing output from the sequencer.
- Proteomics and metabolomics raw data are typically generated by chromatography coupled with mass spectrometry. Raw data are the initially recorded data, such as chromatograms with automatically integrated peak areas and/or mass spectra. For proteomics, several file formats capturing either single runs, paired files, or folders containing several files per run are available. These are often formats from proprietary instruments that are seldom interchangeable and include common formats such as RAW (ThermoFisher, PerkinElmer and Waters), D (Agilent) and WIFF (ABI/SCIEX).

Raw data and quality metrics must be stored in a manner that they cannot be manipulated, and they should be readable for at least the time required by GLP specifications. Ideally, they should be stored as electronic raw data by means of an appropriate validated software (*cf.* Section 2.6). If this is not possible, paper raw data (i.e. immediate printouts with date and signature of the investigator) constitute an acceptable alternative. For ‘traditional’ study types, e.g., using the test methods described in the OECD Test Guidelines, the amount of raw data (and derived data) is often manageable and can be handled entirely, or at least partially, in paper form. By contrast, ‘omics technologies produce huge amounts of initially recorded data. The volume of data in typical ‘omics studies generally makes it difficult if not impossible to record data as paper printouts. Thus, some sort of electronic raw data recording and processing will be necessary. Advancing tools for the GLP-compliant electronic archiving of raw data obtained from ‘omics technologies was one of the challenges discussed at the ECETOC workshop (Buesen et al., 2017). Critical in this context is the overarching problem of how to warrant data integrity. This is an issue that applies to all forms of digital data, be it the original raw data or any processed or archived derivatives (*cf.* respective sections of this article). Checksums provide an elegant solution to this

problem. They provide an algorithm-based state-specific snapshot of the respective data record, with any subsequent change leading to a mismatch. As digital fingerprint checksums will hence reliably indicate data corruption as well as user manipulation.

## 2.5 Data processing and storage

Apart from the amount of raw data, compared to ‘traditional’ study types, there are also usually much more data processing and calculation steps required to translate the data derived from ‘omics technologies into the final interpretable results.

The OECD principles of GLP require that all processing steps are properly documented so that the reported results can be fully reproduced from the raw data. Such a reporting format was the subject of the ECETOC workshop work stream 2 *Towards establishing criteria and best practices for analysing ‘omics-based data*. Workshop participants recommended the development of a standardised, albeit generic and flexible ‘omics reporting framework that could eventually accommodate different ‘omics technologies (Buesen et al., 2017; Gant et al., 2017; in this journal Supplement).

In accordance with the principles of GLP, modifications of records must be justified, and the initial data entry must still be readable. Non-documented data modifications are not permissible. For ‘traditional’ study types, in which observations and results are recorded and processed electronically, this GLP requirement is usually fulfilled by the data recording systems that include an integrated audit trail that documents all data modifications including the respective justification.

When aiming to establish a GLP context for collecting, storing and curating ‘omics data, a special challenge arises from the fact that such an audit trail is generally not available for any of the common electronic evaluation systems. Therefore, changes of ‘omics data have to be documented by other means, at least for the time being. One ‘GLP-like’ solution could be the obligatory requirement to immediately document all data changes or processing steps in a GLP-compliant manner. This could be achieved by an accurate description of the procedure by which the raw data are changed and by storing both the original and the new ‘omics data. The new data should be considered as ‘derived raw data’ and stored with similar requirements as the original raw data. For transcriptomics, this requirement could imply full documentation of the quality assurance and quality control processes implemented in the laboratory; the image analysis software and signal measure used (e.g., processed signal, median, etc.); data normalisation (algorithms or software tools used; test description; a specification if and how confounders were adjusted); filtering (e.g., to remove poor spots or bad arrays) and background adjustments, if applied; any additional filtering (e.g., to remove absent probes from further analysis); and statistical analysis (including algorithms or software tools used).

Further, the documentation procedure has to be audited, and this is one of the tasks of the Quality Assurance Unit (QAU). The QAU is in charge of conducting inspections to verify if GLP studies are performed in compliance with the principles of GLP and if the reported results reflect the raw data accurately and completely. If a QAU has to be established, it should be considered that quality assurance staff must be independent from the laboratory



staff involved in the audited studies. The inspections are described in the Quality Assurance statement included in each GLP report. To fulfil the tasks described above, the QAU must have reading access to all data evaluation systems used. Moreover, it is a prerequisite that the QAU staff has a thorough understanding of the corresponding 'omics technology (e.g., by successfully completing comprehensive trainings) to be able to perform audits and to assess data quality in a profound and reliable manner.

## 2.6 Reporting

'Omics studies have to be reported taking into account the GLP reporting requirements (OECD, 1998). In summary, the final report of an 'omics study should be produced as a detailed scientific document outlining the purpose of the study, describing the methods and materials used, summarising and analysing the generated data, and stating the conclusions drawn from the study. Provided that materials and methods as well as results are well documented, these general GLP reporting requirements could be accomplishable for 'omics studies and 'traditional' studies alike.

Notably, the study director of a GLP study has to sign and date the final report to indicate acceptance of responsibility for the validity of the data and to indicate the extent to which the study complies with the principles of GLP (OECD, 1998). As described above, it may not be possible to conduct all individual components and steps of a study in a fully GLP-compliant manner. In such cases, an 'as GLP-like as possible' procedure (that is governed by GLP guidance) will help to promote the regulatory reliability of 'omics data (e.g. by ensuring transparency on, and confidence in, the applied methodologies and processes). In the GLP compliance statement of the study director, procedures of the study which were not fully GLP compliant (if any) have to be described and assessed with respect to the influence on the study. Thereby, the use of GLP-like procedures can be addressed in a GLP report. Whereas GLP principles are applicable for laboratory processes and data analysis (e.g. statistical methods/software for data analysis have to be validated), they are not applicable to scientific data interpretation (Figure 2). As long as the final study report correctly considers the raw data, the plausible interpretation of data is solely based on the expertise and scientific judgement of the study director. Participants of the ECETOC workshop recognised that performing an 'omics-based study under GLP and/or GLP-like conditions will not necessarily mean that the interpretation of data will be easier.

## 2.7 Software validation

Usually, the application software obtained from a vendor or third party developer (sometimes customised for the corresponding user) is used for 'omics data processing, calculation and interpretation processes. The software used within GLP studies must be validated to ensure reproducibility and traceability of the calculations and correctness of the results. Details concerning computer systems and their validation under GLP are described in the *OECD GLP Advisory Document No. 17 on the application of GLP principles to computerised systems* (OECD, 2016) which also includes aspects of Good Automated Manufacturing Practice (GAMP® 5, 2007).

For ‘omics studies, the software validation efforts might be quite extensive, as ‘omics data evaluation is very complex, and each processing step must be considered. However, software validation can also be achieved by ‘black-box testing’ (also called specification or definition-based testing (FDA, 2002)). Black-box testing implies that test cases are identified based on the definition of *what* the software product is meant to accomplish. The software product can be a complete computer programme or a unit (module) thereof. The test cases are applied to challenge the intended use and functionality of the software product, and its internal and external interfaces. Thereby, the actual algorithms and calculation procedures that together form the application software do not necessarily have to be validated step-by-step. Rather, the process itself is validated, i.e. it is verified that, starting from a given input, the output (i.e. the calculated result) is always the same and that it reflects reality (i.e. the known outcome of the test case). Upon successful completion of the validation, the software product is released. Generally, the validity status of a software product has to be re-checked whenever it is modified. However, the decision, which software modifications require a verification or even full re-validation, depends on a ‘risk analysis’ that the users and/or IT system owners perform, supported by the QAU staff. The state of validation of individual software products may be highly variable, and operators of the ‘omics analysis facility are advised to carefully check coherence with GLP requirements before the onset of the studies.

Sometimes pre-defined Excel spread sheets are in use for calculations. From a GLP point of view, the spread sheet must be validated for correctness of formulas, cell references, etc. This is usually checked by recalculating known results. Additionally, cells with formulas and cell references should be protected to prevent un-authorised or accidental changes, which could affect study results. If it is impracticable to validate a specific Excel spread sheet, e.g. due to a high level of customisation for each individual experiment, the results should be re-checked manually to ensure GLP compliance and reliability.

Finally, for the evaluation and assessment of ‘omics data, separate reference databases might be required (e.g. for pathway analysis of genomics data), and such databases are not necessarily under the validation control of the user. In such cases, the origin of the reference database should be clearly traceable, i.e., at least name, provider/origin as well as version number and/or validity date of the reference database should be described. Moreover, such reference databases could be relevant for the archiving process to ensure reproducibility of data evaluation (*cf.* Section 2.8).

## 2.8 Archiving

As a final step of a GLP study, study plan, raw data and all further study data that are relevant to completely reproduce the study, as well as study report, relevant materials and specimen and the QAU inspection reports have to be archived. The responsible study director should ensure that all relevant materials are archived. The duration of archiving depends on national or agencies’ GLP regulations (e.g., 15 years for paper materials/raw data in Germany).

For ‘omics data, locking and archiving of raw data and processed/derived data is generally only feasible in electronic format. To establish GLP conditions for ‘omics studies, the following rules on archiving should be met:

Data must be reproducible and readable throughout the entire archiving period. Therefore, the data storage media must be controlled regularly for integrity and readability; backups must be kept and controlled, and data storage media replaced where and when considered necessary. Servers or storage media of appropriate capacity should be available, as a huge amount of data might be generated during ‘omics studies.

As mentioned above, also specimens of a GLP study have to be archived. However, this only has to be undertaken as long as the respective sample can be used for evaluation (OECD, 1998). For ‘omics studies, apart from (residual) samples for analyses also specific items, such as DNA chips, could be considered to be specimens. However, such materials are considered to have limited stability to enable proper re-analysis, and specifically required retention periods have not been defined, e.g., in national GLP regulations (as has been undertaken for standard materials (such as histological slides) from ‘traditional’ study types). Hence, the study director of an ‘omics study should assess the stability of specific specimens from his or her study to decide how long archiving would be reasonable, if at all.

In specific ‘omics studies, the data are interpreted using electronic databases containing reference data for many other substances. Such databases can be continuously enlarged and/or refined. Consequently, a specific data interpretation from the past might not be fully reproducible should the reference database change. Such non-reproducibility of study results represents a GLP-compliance issue. To address such scenarios, it would ideally be useful to ‘lock’ or archive, if possible electronically, not only the study data themselves, but also the reference database version used for the corresponding evaluation. This would ensure the reproducibility of calculations and comparisons. However, if this is not possible, other measures (e.g., recording screenshots or version numbers of reference databases) might be more appropriate.

### 3 Workflow of ‘omics studies and consideration of GLP processes: The example of a metabolomics study

This Section exemplarily presents the workflow of a metabolomics study to illustrate the specific components of an ‘omics study and their relevance for GLP. Figure 3 provides a general overview of the workflow of a metabolomics study, beginning with the formulation of a biological question (hypothesis), continuing on to the experimental design of the study, study performance and, finally, the analysis of the data to identify metabolites and the biological interpretation of the data. (Additionally, the Appendix to this article provides a general workflow of studies using RNA sequencing.) Following on from the general overview of the workflow of a metabolomics study, in Figure 4, the general steps of ‘omics studies are juxtaposed with components of the metabolomics study that are relevant for the establishment of GLP conditions.

In the exemplary metabolomics study, animal treatment and blood sampling are performed in a GLP facility, and these steps are covered by corresponding SOPs. For the entire study, the study plan is generated in the animal test facility and a PI at the separate GLP-certified test site is responsible for the analysis of the ‘omics samples (*cf.* Section 2.3 for specific requirements related to multi-site studies). With respect to sample transfer to the analytical

facility (step 1 in Figure 4), care has to be taken that sample transport is undertaken taking into account GLP requirements, such as detailed planning (by study plan and/or SOPs) and documentation, ensuring appropriate environmental conditions during transport and check for completeness and integrity of samples at receipt. Machine-readable labelling (e.g., barcode, radio-frequency identification) can help to avoid any potential mix-ups of samples during transport. The labelling requirements of samples should then be included in SOPs.

Overall, for laboratories that already work in compliance with the principles of GLP, the establishment of GLP conditions should not pose a problem for any of the 4 steps of the metabolomics study, i.e. sample transfer; analytics; data processing; and archiving (Figure 4). The actual analytical measurement itself must be performed using validated equipment and software, and appropriate calibration of the equipment has to be demonstrated. Raw data should reflect the original output of the instruments as closely as possible. For the metabolomics example, raw data are defined as initial analytical data (gas chromatography – mass spectrometry (GC - MS) and liquid chromatography – mass spectrometry (LC - MS) chromatograms, *cf.* step 2 in Figure 4), allowing for a first integration or processing of data if this is performed by a predefined automated algorithm or process, where the operator is unable to influence the output. Taking into account that very large data sets are generally produced during ‘omics studies, the electronic archiving of raw data will be inevitable. As described in Section 2.4 (Definition of raw data), the participants of the ECETOC workshop concluded that this topic should be actively followed up as being analysis-, platform-, and vendor-specific and undergoing rapid development as technological platforms evolve.

Generally, the type of raw data (that have to be defined unequivocally) is different for each type of instrument used. For transcriptomics, the establishment of a GLP(-like) framework is especially challenging due to the rapid advancements of this technology. On the other hand, as new instruments and technologies are developed, the goal to establish GLP(-like) conditions can be brought to the attention of the vendors, and vendors and users should collaborate closely to facilitate the establishment of GLP(-like) conditions. For metabolomics and some proteomics applications, a number of conventionally used analytical instruments (e.g. GC – MS and LC - MS) are already frequently used in a GLP environment. Therefore, the corresponding steps of these studies should be easily amendable to fulfil GLP requirements.

For each step of data processing, a validation procedure is required. As described in Section 2.7 (Software validation), black box testing provides a practical opportunity to largely cover the different steps of ‘omics data processing, as long as they are founded on automated processes using known and archived algorithms. Difficulties may be encountered when data processing requires manual input by the operator or data processing experts, i.e. when the data are processed by non-automated procedures. If it is impossible to avoid manual interactions during ‘omics data processing, a description of the process is required (via SOPs), possibly including criteria for the handling of the data, and the storing of the processed data, as well as corresponding study-related documentation. In the metabolomics example, the processed/derived data are stored in a database called MetaMap@Tox (end of step 3 in Figure 4), in which the data are evaluated by applying statistical methods and by comparing them with reference data comprising, e.g., metabolite patterns from other tested

substances. Apart from the correct and complete data upload process, also the statistical calculations have to be validated (e.g. by comparison with statistical results obtained with the same data sets using different statistical software). The used reference database has to be locked/archived as described in Section 2.8 to ensure retrospective reproducibility of the assessment.

With respect to reporting, clear traceable method descriptions are expected to facilitate the scientific assessment of 'omics data, thereby supporting their applicability in a regulatory context. Importantly, adherence to GLP criteria requires that the final report is reproducible using the original raw data and all documented processes and SOPs. The storing of large amounts of processed data, as derived raw data, again requires that GLP-compliant electronic archiving is readily available. For the metabolomics example (step 4 in Figure 4), all relevant electronic data of a study can be stored on a DVD that is archived in a GLP archive and regularly checked for readability and integrity of data. Notably, at least 2 identical DVDs should be produced for storage in two different locations of a GLP archive, so that a backup copy is always available.

#### **4 Recommendations from the ECETOC workshop related to the quality assurance of 'omics technologies considering GLP requirements**

The following recommendations from the ECETOC workshop aim at facilitating the establishment of GLP(-like) conditions for the handling of 'omics data:

- General standardisation of technologies (e.g., to facilitate common procedures and/or output formats);
- Reduction of human interference in an automated 'all-electronic' environment;
- Checksums of files and storage of the checksums in a non-modifiable way; checksums allow determining whether a file is still in the state it had at the time of computing the checksum or whether it has been modified since;
- Development of a standard ontology in 'omics reporting, to facilitate, e.g., the traceability of conclusions and comparison of assessments.

It was generally recognised that partnering with the vendors of the instruments will considerably enhance the establishment of GLP(-like) conditions for 'omics technologies. Further considering that 'omics, such as transcriptomics, are a rapidly evolving scientific field, partnered instrument development is highly beneficial. Also with respect to the archiving of huge datasets, involvement of vendors is considered valuable to ensure that appropriate data files are obtained. Similarly, vendor-assisted activities to inform and develop an 'omics reporting framework is expected to enhance the transparency and standardised documentation of 'omics studies (Gant et al., 2017; in this journal issue).

Workshop participants highlighted that the training of all personnel working in a GLP environment is an essential aspect of GLP (*cf.* Section 2.1). Understanding the advantages of the principles of GLP for 'omics technologies is expected to increase its acceptance. The

exchange of best practices is generally recommended, particularly when a new framework (the principles of GLP) is beginning to be applied.

A further recommendation from the ECETOC workshop addresses the potential use of elements from the International Standardisation Organisation requirements ISO 17020 and ISO 17025 (ISO, 2005, 2012) to demonstrate accuracy of results in those cases where full GLP status is not yet possible, but scientifically sound processes have to be demonstrated.

Finally, it was recommended to consider the performance-based validation of 'omics technologies to facilitate their regulatory applicability and use. Such performance-based validation can be regarded as an extension of the 'black box testing' to the entire 'omics technology. It includes the use of well-studied (positive and negative) reference substances or appropriate reference samples, if available. If the results obtained for the reference substances or samples are correct, it can be demonstrated that the entire process yields a reliable outcome. Although such performance-based validation may not fulfil standard GLP requirements, it may help to justify that data from 'omics technologies are fit-for-purpose. Workshop participants noted that the scientific goal to use 'omics studies to identify relevant biomarkers of adverse effects is far less reliant on the GLP status of the studies, than the goal to reduce animal toxicity testing in a regulatory context.

## 5 Discussion and Conclusion

When establishing a GLP(-like) environment for 'omics technologies, the most challenging step is most likely the data analysis step of the 'omics study (*cf.* Figure 2). This step includes raw data definition and storage, data processing and software validation as well as appropriate processes to ensure that study results are retrospectively reproducible using the raw data.

The benefit and feasibility to conduct this step of 'omics studies under GLP(-like) conditions largely depends on the goal and type of the respective investigation. If such investigations are performed for research purposes, the required efforts to implement a GLP environment is most likely not counterbalanced by the benefit that the GLP status would promote regulatory acceptance, as such data would only infrequently be submitted to regulatory agencies. For some 'omics technologies, such as transcriptomics/gene expression analysis, it could be quite difficult, or even impossible to achieve full compliance with the principles of GLP, because of the huge amount of data to be processed by validated steps, and to be stored and evaluated, as well as due to the overall difficulties in standardising all of these steps. In such cases, only some steps of the 'omics study can be conducted in a fully GLP-compliant manner, whereas the problematic steps can only be realized in GLP-like manner. To establish GLP-like conditions, those steps of the 'omics studies that cannot be conducted in compliance with the principles of GLP for technical reasons should be clearly identified, just the specific technical reasons that stand in the way to full GLP compliance should be clearly identified.

If it is the study goal to create data-based hypotheses of toxicity or to investigate specific pathways that may be involved in the elicitation of toxicity, the benefit of a GLP(-like)

environment to enhance the use of such data for regulatory hazard and risk assessment may be of secondary relevance since such study results are generally used in weight-of-evidence approaches and not to fulfil toxicological endpoint-specific standard information requirements.

If, however, the ‘omics study is focused on a certain distinct question (e.g., for a transcriptomics study, to prove or disprove the hypothesis that a pre-specified group of genes that can predict a specific toxicological effect are differentially expressed upon substance application), the establishment of a GLP(-like) status is highly relevant to enhance regulatory use of such data. In such cases, the establishment of a GLP(-like) environment may be facilitated by the targeted generation and analysis of reduced amounts of ‘omics data and corresponding reduced process steps, which are possibly also easier to standardise.

For other ‘omics technologies, such as metabolomics, it is less difficult to achieve a GLP status for the data analysis and evaluation steps (*cf.* Section 3). Nevertheless, for a meaningful analysis of the ‘omics data and/or to address specific questions, adequate and comprehensive reference databases are indispensable, and such databases need time to be established. If ‘omics data are intended to directly contribute to the hazard and risk assessment of substances by supporting the assessment of ‘traditional’ toxicity study results (e.g., when a specific metabolite pattern is used to predict a distinct organ-specific effect), the establishment of a GLP(-like) environment will be highly beneficial to obtain consistent high data quality and thus to promote regulatory confidence in the ‘omics data and in the overall assessment.

In conclusion, although a number of challenges remain to obtain full GLP status for ‘omics studies, and particularly for transcriptomics/gene expression studies, the present article describes a clear path forward. Confidence in ‘omics data can be supported by a framework for establishing GLP(-like) conditions for ‘omics studies, as well as by specific complementary activities, as recommended by the participants of the ECETOC workshop. All of these measures aim at ensuring the quality control, reproducibility and traceability of ‘omics studies. This will facilitate their regulatory applicability and use in a fit-for-purpose context.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Abbreviations

<b>ECETOC</b>	European Centre for Ecotoxicology and Toxicology of Chemicals
<b>GC – MS</b>	Gas chromatography – mass spectrometry
<b>GLP</b>	Good laboratory practice
<b>LC – MS</b>	Liquid chromatography – mass spectrometry
<b>NOAEL</b>	No observed adverse effect level

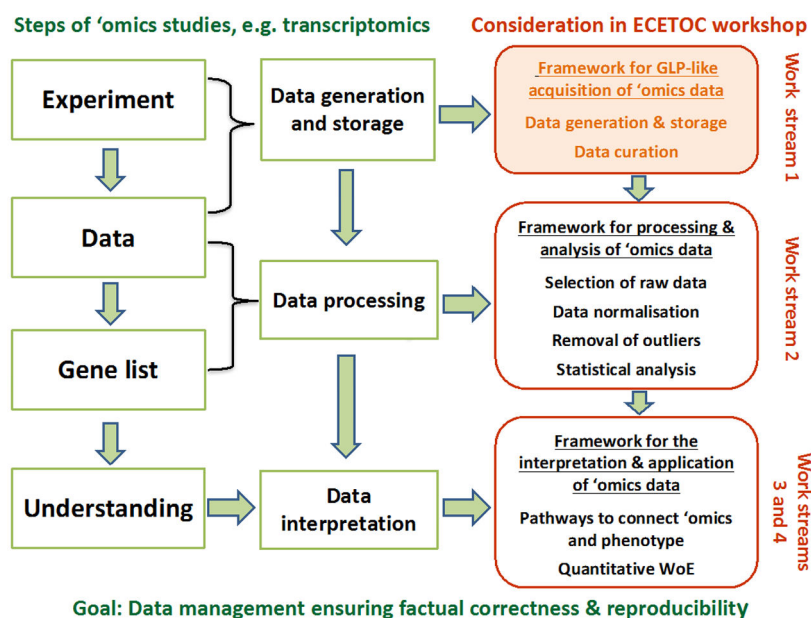
<b>OECD</b>	Organisation for Economic Cooperation and Development
<b>QAU</b>	Quality Assurance Unit
<b>qRT-PCR:</b>	quantitative real-time polymerase chain reactions
<b>SOP</b>	standard operating procedure

## 8 References

- Bridges J, Sauer UG, Buesen R, Deferme L, Tollefsen KE, Tralau T, van Ravenzwaay B, Poole A, Pemberton M, 2017 Framework for the quantitative weight-of-evidence analysis of 'omics data for regulatory purposes. *Regulat. Toxicol. Pharmacol.*, manuscript in preparation.
- Buesen R, Chorley BN, da Silva Lima B, Daston G, Deferme L, Ebbels T, Gant TW, Goetz A, Grealley J, Gribaldo L, Hackermüller J, Hubesch B, Jennen D, Johnson K, Kanno J, Kauffmann H-M, Laffont M, McCullen P, Meehan R, Pemberton M, Perdichizzi S, Piersma AH, Sauer UG, Schmidt K, Seitz H, Sumida K, Tollefsen KE, Tong W, Tralau T, van Ravenzwaay B, Weber R, Worth A, Yauk C, Poole A, 2017 Applying 'omics technologies in chemicals risk assessment: Report of an ECETOC workshop. *Regulat. Toxicol. Pharmacol.*, manuscript in preparation.
- ECHA, 2015 European Chemicals Agency. Read-across assessment framework (RAAF). ECHA-15-R-07-EN, 5 2015, 38 pages.
- EMA (2006). European Medicines Agency ICH Topic S 8 Immunotoxicity Studies for Human Pharmaceuticals, CHMP/167235/2004, 5 2006.
- FDA, 2002 United States Food and Drug Administration. General principles of software validation. Final guidance for industry and FDA staff Issued 11 1 2002. Available at: <http://www.fda.gov/downloads/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/ucm085371.pdf>; accessed 13 January 2017).
- GAMP® 5 (2007). A Risk Based Approach to Compliant GxP Computerised Systems" ISPE Good Automated Manufacturing Practice © ISPE 2007.
- Gant TW, Sauer UG, Chorley BN, Hackermüller J, Perdichizzi S, Tollefsen KE, van Ravenzwaay B, Yauk C, Tong W, Poole A, 2017 A generic Transcriptomics Reporting Framework (TRF) for omics data processing and analysis. *Regulat. Toxicol. Pharmacol.*, manuscript in preparation.
- ISO, 2005 International Standardisation Organisation. General requirements for the competence of testing and calibration laboratories. ISO/IEC 17025:2005.
- ISO, 2012 International Standardisation Organisation. Conformity assessment - Requirements for the operation of various types of bodies performing inspection. ISO/IEC 17020:2012.
- OECD, 1981 Organisation for Economic Co-operation and Development Decision of the Council concerning the mutual acceptance of data in the assessment of chemicals. Paris, France Available at: <http://www.oecd.org/env/ehs/mutualacceptanceofdatamad.htm>; accessed 13 January 2017).
- OECD, 1998 Organisation for Economic Co-operation and Development principles on good laboratory practice (as revised in 1997). OECD principles of good laboratory practice and compliance monitoring. Number 1. ENV/MC/CHEM(98)17.
- OECD, 2002 Organisation for Economic Co-operation and Development Consensus Document No. 13 The application of the OECD principles of GLP to the organisation and management of multi-site studies. ENV/JM/MONO(2002)9. Paris, France, 25 6 2002.
- OECD, 2004 Organisation for Economic Co-operation and Development. Series of good laboratory practice and compliance monitoring No. 14. Advisory document of the working group on good laboratory practice. The application of the principles of GLP to in vitro studies. ENV/JM/MONO(2004)26, Paris, France, 30 11 2004.
- OECD, 2016 Organisation for Economic Co-operation and Development Advisory Document No. 17 Application of GLP principles to computerised systems. ENV/JM/MONO(2002)9. Paris, France, 22 4 2016.

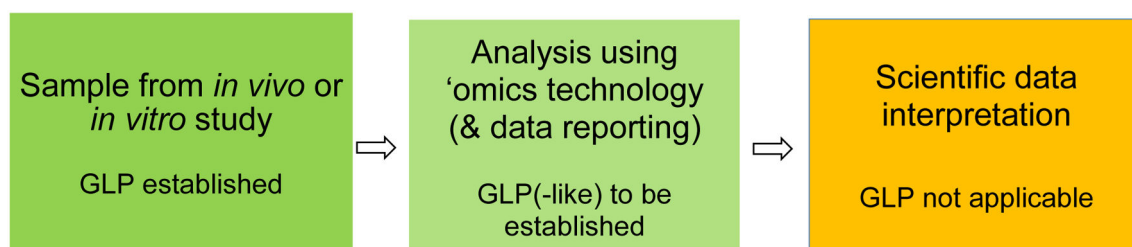


- Sauer UG, Deferme L, Gribaldo L, Hackermüller J, Tralau T, van Ravenzwaay B, Yauk C, Poole A, Tong W, Gant TW, 2017 The challenge of the application of 'omics technologies in chemicals risk assessment: background and outlook. *Regulat. Toxicol. Pharmacol.* manuscript in preparation.
- van Ravenzwaay B, Herold M, Kamp H, Kapp MD, Fabian E, Looser R, Krennrich G, Mellert W, Prokoudine A, Strauss V, Walk T, Wiemer J, 2012 Metabolomics: a tool for early detection of toxicological effects and an opportunity for biology based grouping of chemicals-from QSAR to QBAR. *Mutat. Res* 746(2), 144–50. [PubMed: 22305969]
- van Ravenzwaay B, Sperber S, Lemke O, Fabian E, Faulhammer F, Kamp H, Mellert W, Strauss V, Strigun A, Peter E, Spitzer M, Walk T, 2016 Metabolomics as read-across tool: A case study with phenoxy herbicides. *Regul. Toxicol. Pharmacol* 81, 288–304. [PubMed: 27637788]

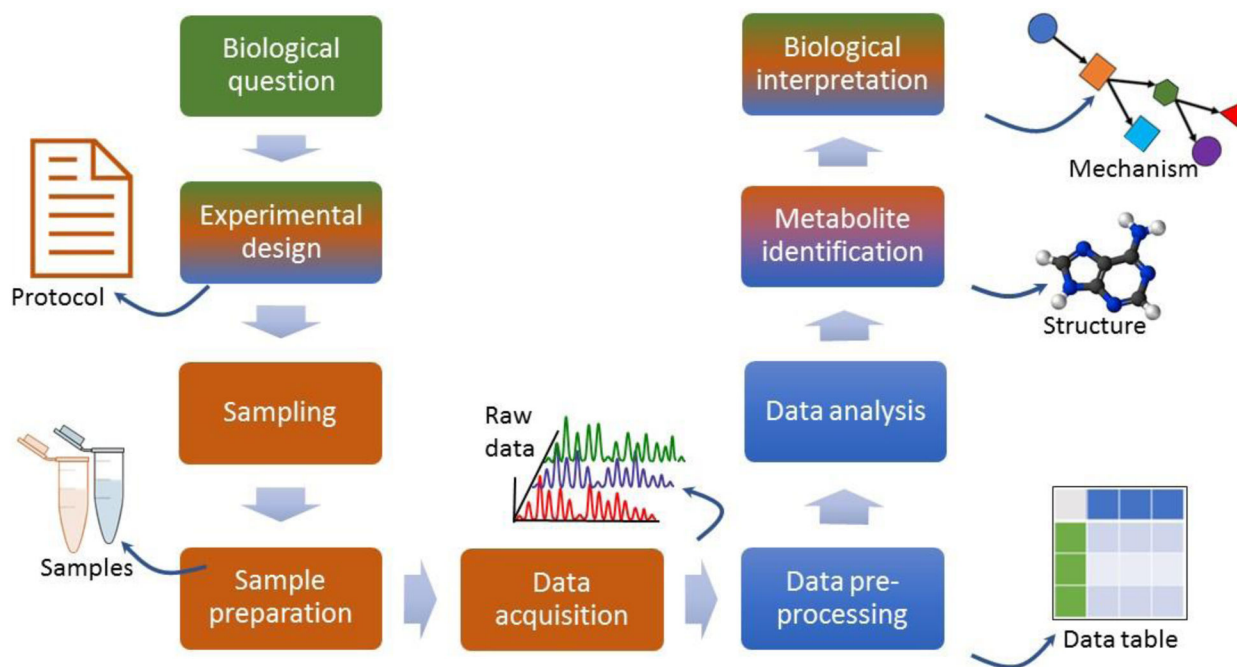


**Figure 1: Overview of the work streams considered at the ECETOC workshop *Applying 'omics technologies in chemical risk assessment* (Buesen et al., 2017)**

**Footnote to Figure 1:** The present article *Quality assurance of 'omics technologies considering GLP requirements* specifically considers discussions from the ECETOC workshop work stream 1 (highlighted in brown).



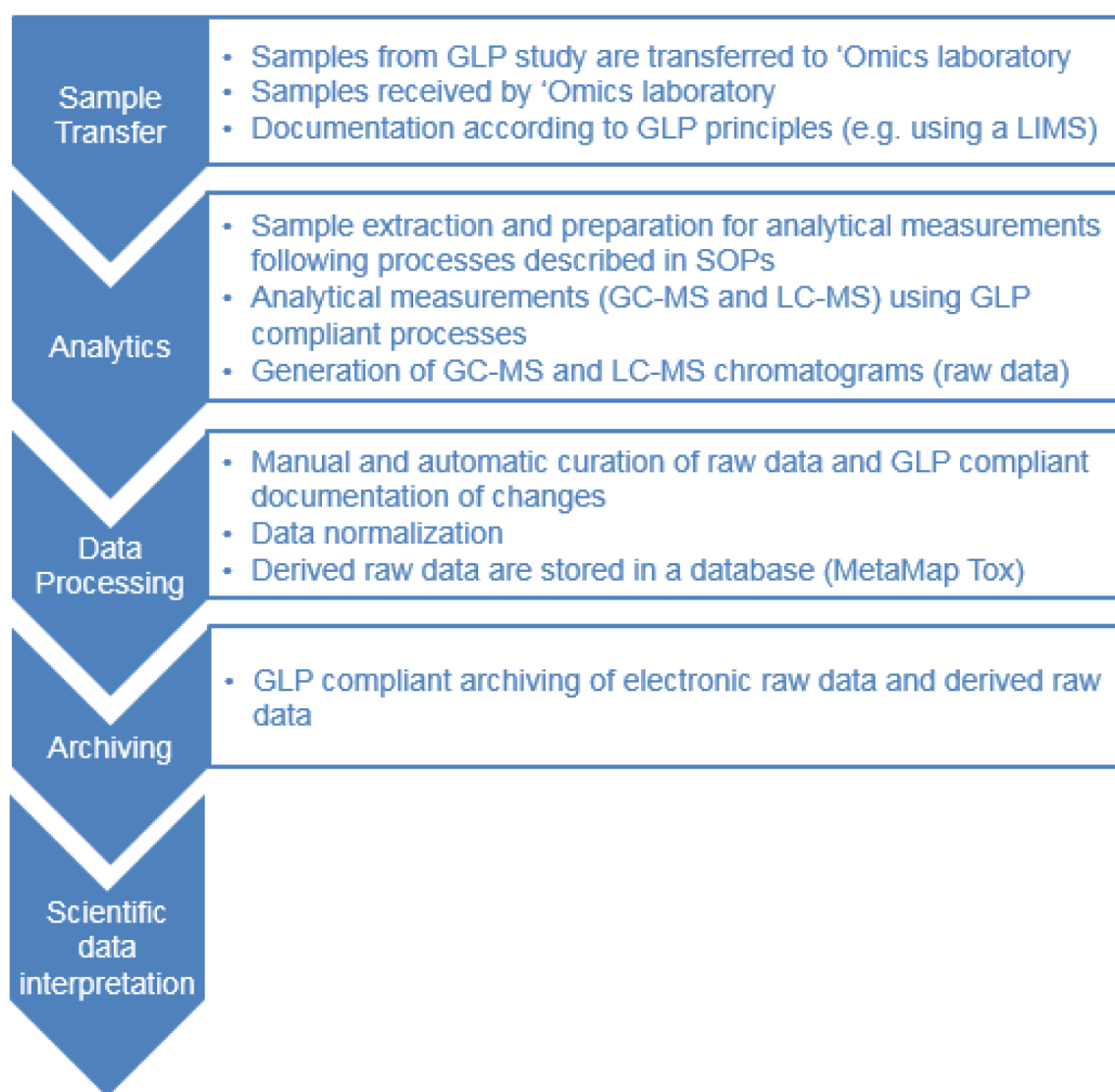
**Figure 2: General steps of an 'omics study with respect to the establishment of GLP(-like) conditions**



**Figure 3: General overview of the workflow of a metabolomics study**

Colour legend to Figure 3: colouring of boxes indicate what expertise is needed at each stage

- green: biologist, orange: analytical chemist, blue: data analyst



**Figure 4: The general steps of 'omics studies, juxtaposed with the workflow of a metabolomics study conducted under GLP(-like) conditions**

Abbreviations to Figure 4: GC-MS: gas chromatography coupled with mass spectrometry, GLP: Good laboratory practice; LC-MS: liquid chromatography coupled with mass spectrometry, LIMS: Laboratory information system; MetaMap® Tox: Metabolomics database (van Ravenzwaay et al., 2012); SOP: Standard operating procedure.