VKGL vereniging klinisch genetische LABORATORIUMDIAGNOSTIEK

VKGL commission for quality_reference standard

Title: Retention Periods	**Doc. code:** VKGL_V05
Subspecialisme: Genome diagnostics	**Version:** 02 **Effective from:** 1-8-2017
Administrator: Els Voorhoeve **Centre:** Erasmus MC, DNA diagnostics	
Specialist explanatory statement on ISO15189:2012 norm 4.13/ enclosure 1 CCKL 4th PRL	

The contents of the 2nd version of this reference standard have been ratified by the VKGL (Professional Association for Clinical Genetics Laboratory Diagnostics) on 18-5-2017, and the VKGN (Professional Association for Dutch Clinical Genetics) on 1-6-2017.

Introduction
The 1st version of this reference standard was ratified by the VKGL (Professional Association for Clinical Genetics Laboratory Diagnostics) on 27-11-2014, and the VKGN (Professional Association for Dutch Clinical Genetics) on 7-10-2014 (per email).
The first version focused primarily on the storage of NGS data, this was due to an urgent need for it in the field. This second version is an extension of the first, with a greater emphasis on the retention periods of various types of data, biological materials, and documents within the genome diagnostics labs. Several retention periods have been adjusted based on a different interpretation of legislation and the feasibility in practice and clarified within a clear schedule.
.

Goal of the reference standard:
Within genetics, there is a need for a national consensus regarding the retention periods of documents, biological materials (samples) and (raw) data. Equally important is what should the laboratories store and for how long?
There are two reasons to want to go back to past results: when a patient without a molecular/cytogenetic diagnosis comes back and would like a re-analysis based on state-of-the-art diagnostics ('state-of-the-art re-analysis'), or when a molecular/cytogenetic diagnosis may have been missed in the past and it should be investigated to see if errors have been made ('cause analysis'). This document contains advice for both scenarios. With 'state-of-the-art re-analysis' it is possible to go all the way back to the source and re-sequence the DNA sample. In the 'cause analysis' the results are subject to mandatory re-assessment.

Scope:
This document applies to all genomic laboratories within the VKGL and primarily concerns the retention of data and materials from advice seekers/patients submitted for genome diagnostics.

Principles:
Background information that has played a role in determining retention periods within clinical genetics:

- Showing a DNA variant or constitutional chromosome change is not a snapshot, but a result for life.
- An incorrect DNA/chromosome result can have very large implications and any mistake that is made, only comes to light years later, e.g., in the case of pre-symptomatic diagnostics.

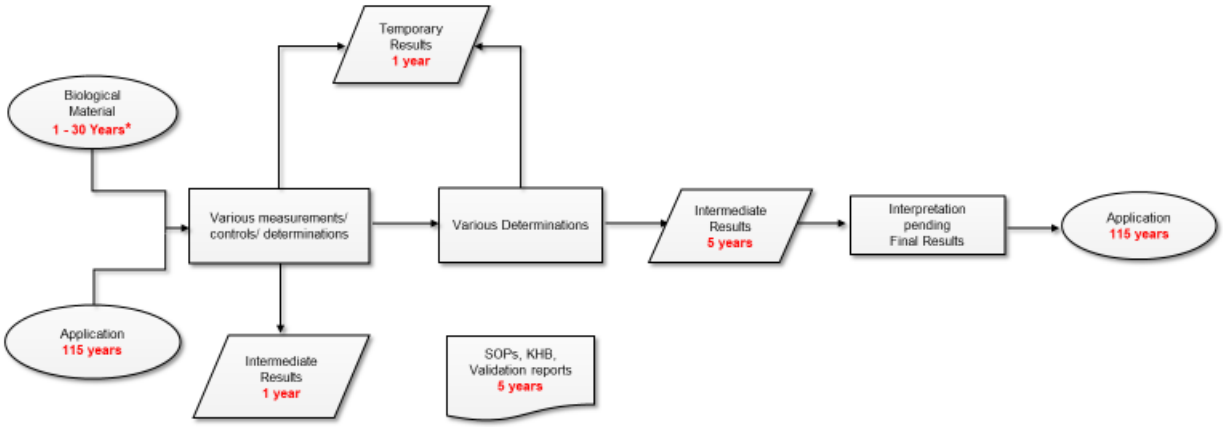
- The cost/benefit consideration for examples such as the storage of very large digital files as those needed for next-generation sequencing (NGS).
  - Based on several publications (see references), a selection has been made of the various sorts of data that we store. The real raw data files are too large to be stored for a longer period of time. The proposal is therefore to only contain retention periods for the various edited data files. The purpose of storing these files is to guarantee the interpretation of the variants at the time of reporting. If there is a possibility of a missed variant, somewhere in the process, then the data analysis must be able to be repeated ('cause analysis'). This reference standard is based on the assumption that the 'cause re-analysis' starts with filtering, for which a variant list is required. The variant calling falls outside the scope of this directive based on the assumption that the pipeline performing the variant calling has been validated for the relevant application. For this reason, it is therefore not mandatory to keep the files that are necessary for the variant calling for a longer period of time. The quality of the NGS run must be clear with regard to coverage of the exome / genome / genes (horizontal coverage) and the number of reads per locus (vertical coverage) (statistics). The specifications of the pipeline used at that time must also be traceable. The retention periods start from the moment the result is reported. In the period between the use of the run and the analysis of the data, all data files must be available and remain that way. The storage of all this data does not have the explicit intention of finding out a possible sample mix-up. Each lab must exclude the sample mix-ups at the time of reporting using internal procedures, such as an SNP check.
- The cost/benefit consideration for keeping outdated software operational in order to be able to continue reading old files in the future.
- Storing results of techniques that have not been operational for a long time has little added value, if, for example, traceability and required knowledge are no longer available.
- Retention periods must comply with all current laws and regulations.
- As long as the DNA is stored, a test can always be repeated.

Definitions and explanations:
Biological material: this concerns both the acknowledged biological material and the processed biological material.
Application: this concerns both a digital application (via a LIMS/hospital system) and a completed application form on paper. An applicant uses the form to request the lab to answer a specific question on certain (submitted) material.
Intermediate results: data from all kinds of measurements / determinations that define the quality and progress of the research but have no direct influence on the result of this research. An example could be measurement of DNA concentration, photos of test gels. For NGS these are: QC values / statistics regarding the quality of the data, percentages horizontal and vertical coverage, FASTQ, XSQ, SAM or BAM (alignment) files, along with information about algorithms / computer science being used.
Final results: data that was directly responsible for the result. Good examples are the variant list (VCF file), print sequence file, karyogram, array profile, autoradiogram Southern blot, and DGGE image.

VKGL vereniging klinisch genetische LABORATORIUMDIAGNOSTIEK	
VKGL commission for quality_reference standard	
Title: Retention Periods	**Doc. code:** VKGL_V05
Subspecialisme: Genome diagnostics	**Version:** 02 **Effective from:** 1-8-2017
Administrator: Els Voorhoeve **Centre:** Erasmus MC, DNA diagnostics	
Specialist explanatory statement on ISO15189:2012 norm 4.13/ enclosure 1 CCKL 4th PRL	

Result letter: report from the laboratory to the applicant containing the results including the interpretation of the requested study.
Quality documents: various documents that are required to carry out the requested research under accreditation and in accordance with the quality requirements that are imposed (such as ISO 15189/CCKL). Good examples are the quality manual, SOPs, validation reports, audit reports, deployment forms and various. tests, report management reviews, proficiency tests, and diverse registrations.
Temporary results: Results that are only generated once and immediately processed. A good example is the raw data for processing by algorithms: bee. BAM, SAM and FASTQ files.

Schedule for periods of retention:



115 Years: this period of time is derived from the addition to the 1985 law on archiving: Base document for public and special Academic Hospitals
30 Years: minimum period of time derived from the preliminary draft of the Health Consumer Rights Law (WCZ), also adopted by the NVVP (Pathology)
5 Years: during this period of time all results and data remain available within an audit cycle for scrutiny of traceability
1 Year: securing the interpretation of results pending the reporting of the Final Results
*: Please see list of specifications for patient materials and retention periods

Periods of retention for biological material:

DNA	30 years
Liquor, urine, plasma*	5 years
Suspensions	1 year
Cell lines	30 years
Tissue	5 years
Blood**	2 weeks to 5 years

VKGL vereniging klinisch genetische LABORATORIUMDIAGNOSTIEK	
VKGL commission for quality_reference standard	
Title: Retention Periods	**Doc. code:** VKGL_V05
Subspecialisme: Genome diagnostics	**Version:** 02 **Effective from:** 1-8-2017
Administrator: Els Voorhoeve **Centre:** Erasmus MC, DNA diagnostics	
Specialist explanatory statement on ISO15189:2012 norm 4.13/ enclosure 1 CCKL 4th PRL	

These are minimum retention periods. A laboratory may decide to keep biological material for a longer period of time, in which case, it is advisable to provide justification for this extended retention period. This should not only consider the 'provision of good care', but also the privacy interests of the patients whose data and biological material are stored. (See also point 7 of the medical lawyer's note at the end of this document).
*: for the plasma of NIPT studies, a much shorter retention period applies. This material should be stored for at least 3 months after delivery.
**: the minimum retention period depends on the intended purpose and the parameters that may still need to be determined. This period is much shorter for haematological parameters, than when the blood is stored in order to isolate DNA at a later stage.

Abbreviations and explanations NGS terminology:

NGS: next-generation sequencing
Gene package: a random package with an average content of roughly 40 genes.
Variants list (e.g., VCF): This is usually a text file containing information and statistics for each variant found. It also contains 'metainformation' (starting with ##) with information about each individual column and usually has information about how the file originated (which software was used, for example). Average size of gene package/exome: ~0.5 Mb/~35 Mb.
BAM: Binary version of the SAM file. This version is usually saved. Average size of the final gene package/exome BAM: ~0.2- 1 Gb/~10-20 Gb.
SAM: Sequence Alignment/Map. This is a tab-delimited text file (often containing a header with information) in which the alignments that have been made are described. Per read, a unique name, chromosome, location on chromosome, certain quality values, if present: pair information and there are many optional values that can be provided (whether a read is uniquely mapped or not mapped, for example). Since the SAM file is very large, it is converted directly to the binary version (BAM), because it is much smaller, and most programs can accept both BAM and SAM as input.
FASTQ/ XSQ: This is the raw data file that comes from the sequencer. FASTQ is not an abbreviation, but a contraction of FASTA and Quality. FASTA is a text file that contains sequence information (for example, reads created by the sequencer). A FASTQ file also contains quality information. Average size of gene packet/exome FASTQ files depends on how many samples are loaded per lane. For exome between 4 and 20 Gb per sample/gene package is more stable with about 4 Gb per sample (zipped ~250 Mb).

VKGL vereniging klinisch genetische LABORATORIUMDIAGNOSTIEK	
VKGL commission for quality_reference standard	
Title: Retention Periods	**Doc. code:** VKGL_V05
Subspecialisme: Genome diagnostics	**Version:** 02 **Effective from:** 1-8-2017
Administrator: Els Voorhoeve **Centre:** Erasmus MC, DNA diagnostics	
Specialist explanatory statement on ISO15189:2012 norm 4.13/ enclosure 1 CCKL 4th PRL	

Legal Note:

Legal Note regarding retention periods genetic data and biological material
M.C. Ploem, 6 december 2013

1. The Personal Data Protection Act (General Privacy Act) contains the basic standard for the retention of personal data: storage is permitted as long as this is necessary for the purpose for which personal data have been collected and stored; if the necessity of storing personal data ceases, the identification for this also ceases; the data may be stored in anonymized form, which means that the patient behind the data can no longer be traced (this is not the case with so-called 'encrypted' data).

*2. A special arrangement applies to the retention of **patient** data, which is laid down in the WGBO. In short, this regulation means that all data that are relevant for the proper treatment of the patient are stored for **15 years** or **as long as results from the care of a good care provider** (Art. 7:454 paragraph 3 DCC). This obligation is also known as the 'file obligation'.*

3. In the event that, following the performance of a (genetic) test in the lab, data are released that do not fall under the doctor's obligation to file (because they are not relevant for the treatment of the patient), the retention period of (at least) 15 years is not applicable. Retention of the relevant data may nevertheless be necessary, e.g., for the validation of the test carried out and/or quality monitoring within the lab. Such data may be retained for as long as necessary in view of those purposes (quality and safety). I assume that this is a period of 1-5 years after the test has been carried out. If a researcher would like to store such lab data and use it for future scientific research, the patient must be carefully informed about this (including information on the likelihood of 'new findings' and policy in this regard) and have given explicit permission (unless the data is anonymized; see above).

4. The storage of biological material is not regulated in the WGBO. However, it can be deduced from general standards (Article 10 of the Constitution, Personal Data Protection Act) that this may/must be done for as long as necessary to provide good care for the patient (and if the material is used for research: as long as necessary to be able to do good research; the retention of material for research is primarily linked to the explicit consent of the patient).

*5. For (the public) academic hospitals, different provisions apply on a number of points. These are related to provisions that are made by or pursuant to the Archives Act. It follows from a 'Basic Document public and special academic hospitals 1985' based on that law that some data from the patient file are kept for **115 years**, calculated from the birth of the patient. This concerns: discharge letter, operation report, anesthesia report, PA report (histology, cytology and /or necropsy), first aid report and documents containing information about calamities. An*

important question is to what extent this also includes genetic data. It seems to me that certainly not all data generated by a genetic test fall under the concept of 'core data'. The result of a specific genetic test (e.g. that a newborn baby has sickle cell disease) may be, but the raw test data probably is not.

6. The long-term availability of patient data and biological material has various consequences for the patient. It can obviously serve the medical interest of the patient, but also harm his personal or social interests, for example when his data and/or material come into the hands of investigating officers or unauthorized persons or are requested when entering into work or for insurance purposes. One can also think of the psychological burden on the person concerned if analysis of his material and/or data shows that there is/is an increased risk of a serious not (easily) treatable disease. These risks are particularly relevant when it comes to a person's exome/genome data or biological material, because not only can a person's identity be deduced (increasingly easily), but also which diseases a person has, or which health risks the person concerned runs (in the long term).

*7. The following can be derived from the foregoing. The position that has been taken that good care within the genetics standard (and not only in an individual case) entails a longer retention period (than 15 years) of the relevant care data (and body materials) seems to me to be entirely defensible. How much longer than 15 years that period of time should be, is primarily a question that the professional groups involved (medical specialists, in particularly geneticists in consultation with employees of the lab) **must answer together**. In doing so, not only must account be taken of (an adequate interpretation of the standard of) 'the provision of good care', but also of the privacy interests of the patients whose data and material are stored. This means, for example, that for the very long-term (115 years) storage of the most privacy-sensitive data, such as a DNA code (or biological material that is undoubtedly sensitive to privacy), a very good reason must be present in the medical file, and that this must be enclosed by the best possible privacy guarantees.*

References:

NGS

- Sboner, A., Mu, X.J., Greenbaum, D, Auerbach, R.K., & Gerstein, M.B. The real cost of sequencing: higher than you think! Genome Biol. 12, doi:10.1186/gb-2011-12-8-125 (2011).
- Weiss et al. Best Practice Guidelines for the Use of Next-Generation Sequencing Applications in Genome Diagnostics: A National Collaborative Study of Dutch Genome Diagnostic Laboratories. Hum Mutat. 2013 Jun 17. doi: 10.1002/humu.22368.
- Practice guidelines for Targeted Next Generation Sequencing Analysis and Interpretation. Van de CMGS website: http://cmgsweb.shared.hosting.zen.co.uk/BPGs/BPG%20for%20targeted%20next%20generation%20sequencing%20final.pdf
- http://pathwiki.rcpaqap.com.au/pathwiki/index.php/IT_infrastructure#Data_management_and_storage

VKGL vereniging klinisch genetische LABORATORIUMDIAGNOSTIEK	
VKGL commission for quality_reference standard	
Title: Retention Periods	**Doc. code:** VKGL_V05
Subspecialisme: Genome diagnostics	**Version:** 02 **Effective from:** 1-8-2017
Administrator: Els Voorhoeve **Centre:** Erasmus MC, DNA diagnostics	
Specialist explanatory statement on ISO15189:2012 norm 4.13/ enclosure 1 CCKL 4th PRL	

Others

- http://www.federa.org/2007-bewaren In dit stuk een pleidooi voor het langer bewaren van medische dossiers in de klinische genetica en de pathologie, mede gezien het belang voor familieleden. Een periode van 100 jaar wordt genoemd. Er is nog geen nieuwe landelijk regelgeving.
- http://wetten.overheid.nl/BWBR0014594/geldigheidsdatum_27-09-2013#Opschrift. Planningsbesluit klinisch genetisch onderzoek en erfelijkheidsadvisering. In paragraaf 3.2 aandacht voor landelijke afspraken voor het bewaren van lichaamsmateriaal.
- Huidige regelgeving: 15 jaar bewaren medische dossiers.
- Archiefwet: UMC's moeten basisdocumenten van het dossier 115 jaar bewaren. Basisdocumenten zijn in ieder geval: aanvraag, uitslagbrief.
- Cytogenetische richtlijn van de EUROPEAN CYTOGENETICISTS ASSOCIATION (ECA): http://www.e-c-a.eu/files/downloads/Guidelines/E.C.A._General_Guidelines_Version-2.0.pdf. Hierin de bewaartermijnen voor cytogenetica, voor gevallen die mogelijk niet via nationale wetten/regelgeving zijn geregeld. Hierin minimaal bewaartermijn van 5 jaar, maar liefst oneindig indien er een afwijking is gevonden. Hiermee worden vooral de cellen (stikstof) en de karyogrammen (digitaal of hard copy) bedoeld.
- http://www.gezondheidsraad.nl/sites/default/files/04@08n1.pdf rapport van de gezondheidsraad aan minister uit 2004. Hierin pleidooi voor het langer bewaren dan de huidige 10/15 jaar.
- *Requirements for the Retention of Laboratory Records and Diagnostic Material* (2009) NPAAC, Australische richtlijn voor pathologie, revisie 2013. Zie paragraaf minimum retention times genetics, table 5: http://www.health.gov.au/internet/main/publishing.nsf/Content/B8562E2C3D131ED8CA257BF00019153C/$File/V0.24%20Retention.pdf
- https://www.acmg.net/StaticContent/SGs/Section_E_2011.pdf. Standards and guidelines for clinical genetics Laboratories van American College of Medical Genetics.
- Advies bewaartermijnen van de Nederlandse Vereniging voor Pathologie. Zie http://www.pathology.nl/images/actueel/Publicaties/Kwaliteit/Adviesbewaartermijnendefinitief_juli_2010.pdf