Research documentation at Karolinska Institutet

A handbook

Table of Contents

Research at KI

All research at Karolinska Institutet (KI) should be carried out in the best way and with the highest quality possible. It must always be conducted according to national and international laws and regulations to ensure safe and correct procedures. This is especially important when the research involves humans and other animals.

Research should be carried out according to general rules of good research practice[1]:

- **Integrity** – researchers should work with honesty and respect and strive to eliminate fraud and misconduct.
- **Openness** – research should be as open as possible (without violating individual integrity, intellectual property, laws and regulations).
- **Professional guidance and legislation** – all researchers must be aware of the legal requirements governing their research and general rules and guidelines within their respective field.
- **Leadership, co-operation, supervision and training** – should be freely accessible to all researchers. All PhD students have the right to supervision and appropriate training should be provided.
- **Primary data/samples/equipment** – ownership and use should be clarified from the onset (especially important in collaborative projects), both for the current project and for possible future research. Proper documentation of the research is essential.
- **Dissemination and publication of results** – not just within the academic world but also to the general public (the latter is referred to as the "third assignment" [tredje uppgiften] in Sweden, following the first two assignments covering research and education).
- **Ethical practice in research involving humans and other animals** – rules and regulations must be followed. This includes necessary approval from the relevant ethical board or committee before the commencement of the research project.

More detailed information is available on the internet, for example at the Swedish Research Council's[2] (Vetenskapsrådet, VR) website. Among other things, they have published a report on good research practice[3] and have also compiled information about rules and guidelines for research in their codex[4].

[1] See for example: KI's previous "Guidelines for planning, conducting and documenting research" (November 2006, Dnr 4820/06-600), Rules of Good Scientific Practice (Max Plank Society, November 2000, http://www.mpg.de/pdf/rulesScientificPract.pdf), Good Research Practice (University of Cambridge, 2008, http://www.rsd.cam.ac.uk/research/Good_Practice.aspx), Guidelines on Good Research Practice (Wellcome Trust, November 2005, http://www.wellcome.ac.uk/About-us/Policy/Policy-and-position-statements/WTD002753.htm).

[2] Swedish Research Council = Vetenskapsrådet, abbreviated VR (www.vr.se)

[3] Good research practice – what is it? Views, guidelines and examples; Bengt Gustafsson, Göran Hermerén & Bo Petersson (http://www.vr.se/download/18.6b2f98a910b3e260ae28000469/Good%2BResearch%2BPractice.pdf)

[4] CODEX – rules and guidelines for research (http://www.codex.vr.se/en/index.shtml)

Information on ethical rules and regulations specific to KI can be found on the Ethical Council's webpage[5]. Issues addressed on their webpage include, among others: "Ethical guidelines for international collaborations"; "Rules governing conflict of interest"; and "Ethical policy for supervision of doctoral students from other countries and cultures". Any researcher at KI may contact the Ethical Council for advice on ethically related research questions.

Why document your research?

There are several reasons why research must be documented in a high quality and intelligible way:

- It should be possible to go back and repeat an experiment/test/analysis/etc. and arrive at the same result/conclusion again.
- A good recording of a procedure/protocol or steps taken will facilitate troubleshooting.
- If there is any suspicion of fraud, researchers must be able to provide proof (usually in the form of documentation) of what has been done and that it was done in the way claimed. Good quality documentation will help to protect against fraud (and misconduct).
- The overall quality of the research improves with clear and detailed documentation.

Regardless of the field of research each new project and its experimental outline should be written in a project plan, which should also include:

- Project title
- Specific aim(s)
- Patients/volunteers studied (when relevant)
- Experimental material
- Procedures and measurements
- Data collection and calculations
- Planned statistical analyses

For clinical and epidemiological research the project plan and documentation may also include protocols, questionnaires, case report forms (CRF:s) and the like. These should also be clear and unambiguous for both co-workers and outsiders.

Laboratory notebooks (log books, journals) must be maintained – either on paper or in an electronic form – with entries made for all experiments, observations and/or measurements. Any alterations must be dated and signed, and entered in such a way that the original text is still legible (especially important when using paper laboratory notebooks). The information should be kept for 10-15 years depending on the kind of research, see a later section on archiving demands for more information.

As far as possible, results must be documented and retained in the form of original/raw data, in a legible, easily identifiable and readily accessible form, for example:

- Permanently stored electronically or permanent printouts of electronic data
- Printouts from measuring instruments or an attached printer (if not stored electronically)

[5] KI's Ethical Council = Etikrådet (http://intra.ki.se/organization/committees/Ethical_Committee/index_en.html)

- Photographs of morphological data or molecular biological analyses (e.g. gels) which can be indentified with certainty
- Printouts of statistical calculations (if not stored electronically)

Unpublished results should be kept also, regardless of research field and results of interviews or questionnaires and information from/in patient case notes must also be filed securely.

A main rule is that it must be possible to identify the original observations/results on which published tables and figures are based. This is important for legal and ethical purposes during reviews, follow-ups, etc. and whenever comparisons amongst results must be made.

Rules and regulations affecting research at KI

While doing research at KI you need to follow Swedish legislation and the local rules and regulations at KI (these also comply with international regulations). Depending on the kind of research you conduct, a combination of various laws may apply. For research that in any way involves humans or animals, an ethical approval is needed before the research begins. For research involving humans, ethical approval is obtained from the ***Ethical Review Board*** (Etikprövningsnämnd) and for research involving animals, it is granted by the ***Ethical Committee on Animal Experiments*** (Djurförsöksetisknämnd).

Approvals may also be required from the Radiation Protection Committee (Strålskyddskommittén), the Medical Products Agency (Läkemedelsverket) or the National Board of Health and Welfare (Socialstyrelsen).

International legal instruments that regulate human research are, for example:

- The Nuremberg and Helsinki Declarations[6]
- The European Convention on Human rights and Biomedicine[7]

Swedish laws for research involving humans:

- **The Ethical Review Act**[8] (Lag om etikprövning av forskning som avser människor)
- **The Personal Data Act**[9] (Personuppgiftslag, abbreviated PUL)
- **The Public Access to Information and Secrecy Act**[10] (Offentlighets- och sekretesslag)
- **The Biobanks in Medical Care Act**[11] (Lag om biobanker i hälso- och sjukvården m.m., also known as Biobank Law or Biobankslagen)

Swedish law for laboratory animals:

- The Animal Welfare Act[12] (Djurskyddslag)

Even if you do not need ethical or other approval for your research, you still need to follow the regulations for research documentation and archiving.

---

[6] The Nuremberg Declaration on Peace and Justice (http://www.peace-justice-conference.info/declaration.asp) and the Helsinki Declaration on Ethical Principles for Medical Research Involving Human Subjects (http://www.wma.net/e/policy/pdf/17c.pdf)

[7] The European Convention on Human rights and Biomedicine, CETS No 164, http://conventions.coe.int/Treaty/Commun/QueVoulezVous.asp?NT=164&CL=ENG

[8] The Ethical Review Act = Lag om etikprövning av forskning som avser människor (2003:460)

[9] Personal Data Act = Personuppgiftslag (1998:204)

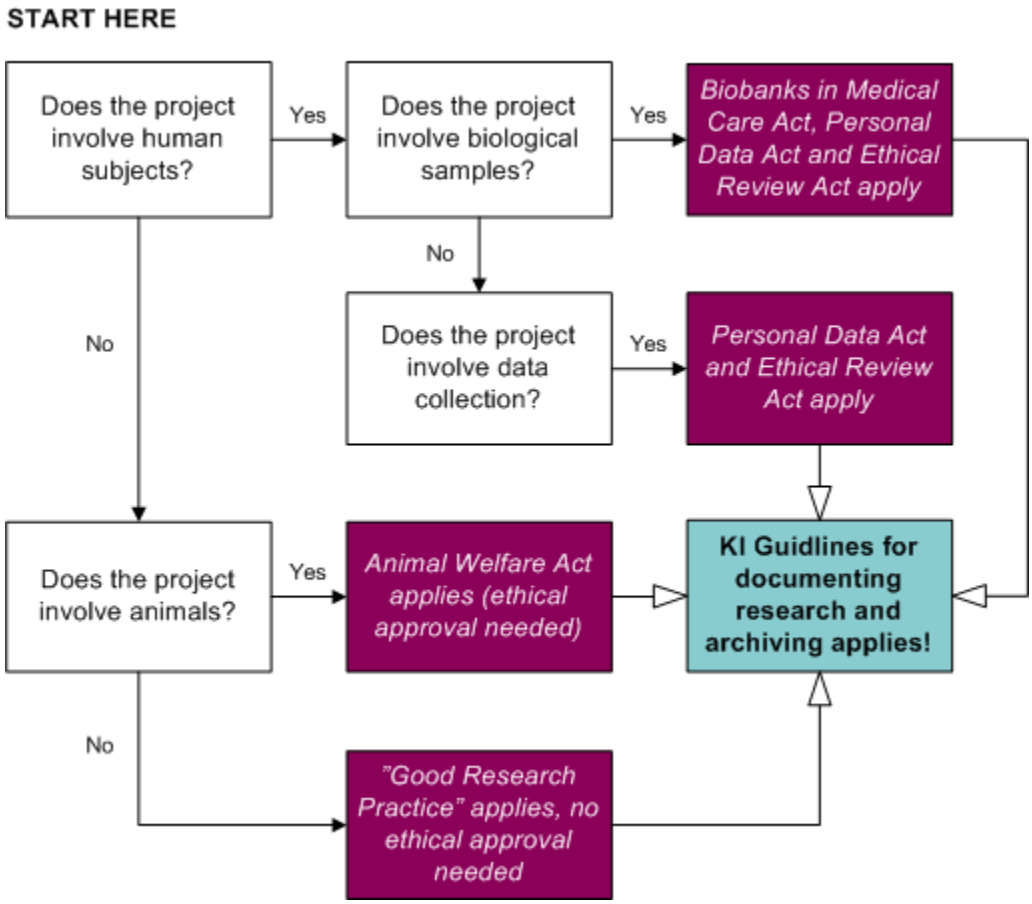
[10] Public Access to Information and Secrecy Act = Offentlighets- och sekretesslag (2009:400)

[11] Biobanks in Medical Care Act = Lag om biobanker i hälso- och sjukvården m.m. (2002:297)

[12] Animal Welfare Act = Djurskyddslag (1988:534)

How do you know which law(s) apply to your research?

Use the flowchart below to help you find out. Below the flowchart will follow short descriptions of the main points from Swedish legislation and KI guidelines and, when applicable, contact information will be provided. **Keep in mind, however, that each individual researcher is responsible for making sure that he or she has the necessary information regarding rules and regulations affecting his or her own research.** Sometimes the information given below will not be sufficient and you will have to go to the original texts, webpages, etc. for complete information.



The Ethical Review Act

In 2003 the **Ethical Review Act** (Lag om etikprövning av forskning som avser människor, also known as Etikprövningslagen) came into force in Sweden. The Ethical Review Act regulates research on humans or involving human biological material. An ethical approval is needed before the research may begin. Ethical vetting of proposed research involving humans is done by one central and six different ***Regional Ethical Review Boards*** (regional Etikprövningsnämnd, EPN). If consensus cannot be reached within a Regional Ethical Vetting Board, or if a researcher wishes to appeal the decision of a regional board, the final decision may be referred to the

Central Ethical Review Board. Visit the EPN webpage for more information (http://www.epn.se/start/startpage.aspx[14]).

The Personal Data Act

The **Personal Data Act** (Personuppgiftslagen, also known as PUL) regulates handling of personal data, with the purpose of protecting against personal integrity violations. The law came into force in 1998.

What is personal data?

Personal data is information that can be linked directly or indirectly to a specific person alive today – it could be everything from a personal identification number (personnummer) to a vehicle registration number. It is divided into sensitive and non-sensitive data, where sensitive data is specified as information about race and ethnic origin, political opinions, religious or philosophical beliefs, union affiliation, health and sexual life. Personal data needs to be handled in a secure way at all times. This means, for example, that if the information is kept on "detachable media" (USB memory sticks, compact discs, laptops etc.) it should be encrypted and must be handled with care. All personal data must be kept safe from unauthorised access. Information kept in the form of regular paper documents, e.g. questionnaires, must be handled in a secure way according to specific rules and should be kept locked.

If you as a researcher plan to compile personal data into a registry, this registry must be reported to the ***Personal data representative*** (Personuppgiftsombudet) at KI Victoria Söderqvist (victoria.soderqvist@ki.se, 08-524 864 73).

Processing genetic information

If the research includes linking personal data together with genetical analysis the ***Swedish Data Inspection Board***[15] (Datainspektionen) needs to be notified no later than three weeks prior to the start of the processing. An application form can be downloaded from the Swedish Data Inspection Board's webpage (www.datainspektionen.se).

The right to access public documents in Sweden (*Offentlighetsprincipen*)

Occasionally, news reports arise about researchers refusing to hand out information regarding their research, arguing that they have promised their research subjects not to share information. Unfortunately, there is no legal basis for such a promise. In Sweden, everyone has the right to study documents held by public authorities. As a Swedish university (and thereby a public

[13] Ethical Review Board = Etikprövningsnämnd, http://www.epn.se/start/startpage.aspx

[14] Ethical Review Board = Etikprövningsnämnd, http://www.epn.se/start/startpage.aspx

[15] Application for genetical analysis to the Swedish Data Inspection Board = Blankett för föranmälan av genetiska analyser till Datainspektionen, http://www.datainspektionen.se/Documents/blankett-forhandskontroll_genetisk.pdf

authority), KI falls under the same legal obligation of public access. However, certain limitations exist on the public's right to access such information:

- The public access principle applies only to documents classified as being public.
- A public document can be covered by secrecy, in which case some restrictions apply.

A document in this meaning is every arrangement that is communicating a message, that is, not only paper documents, but also an e-mail, for example. A blood sample is not seen as a document in spite of the fact that it contains a lot of genetic information traceable to the donor; however, a report on the sample(s) is considered a document in this regard. A public document (allmän handling) is any information that is received, dispatched, drawn up and stored at KI. Anyone may request access to them. When granted access, relevant legislation will apply to the recipient of the information.

It is very important to know that a request to obtain a public document must be dealt with immediately and released without unnecessary delay. This usually means that if someone wants to see a document within your research, any current work should be paused to help that person (if this can be done without negatively affecting your work). More details can be found in **the Freedom of the Press Act**[16].

Examples of documents that are not considered public are personal notes or memoranda, personal letters and working material. ***Working material*** (arbetsmaterial), which is normally not regarded as a public document, becomes such if it is officially archived at KI. For examples and definitions of working material, see the KI **Archiving plan**[17].

Ultimately, KI is responsible for all research conducted at the university and is the legal "owner" of the raw/primary data. ***Lärarundantaget*** (the teachers' exception[18]) gives researchers at Swedish universities and higher education facilities the right to their own results, but not their own raw data. The researcher's right to his/her own results and thoughts are of importance for intellectual property (IP) issues and prospective patents.

Public Access and Secrecy

Information concerning individuals participating in a research study is protected by the **Public Access to Information and Secrecy Act** (Offentlighets – och sekretesslag), but the secrecy is not absolute. Generally, it applies to information that cannot be revealed unless it is obvious that the individual in question or someone close to him/her will not suffer from the revelation. Access to the information can be requested and granted for other research and statistics. It is very important, therefore, never to promise research subjects absolute secrecy, but instead clearly state that information might be given out. Coded or de-identified information may also be given out if requested. Coded information means that the original persons can be traced using a code

[16]The Freedom of the Press Act = Tryckfrihetsförordningen (SFS 1949:105)

[17]Archiving Plan = Arkivbildningsplan för institutionerna vid Karolinska Institutet (Dnr: 4159/08-040)

[18] Teachers' exception = Lärarundantaget, more information can be found on http://ki.se/ki/jsp/polopoly.jsp?d=3620&a=10115&l=en

key, while in the case of de-identified (de-personalised) information, the code key has been destroyed.

Secrecy regulations also apply to research financed by another principal entity (huvudman) or organisation, so called commissioned or sponsored research (uppdragsforskning), or some cases of animal based research. This is most often relevant with regards to future or pending patent applications.

Information protected by the Public Access to Information and Secrecy Act must be handled with care, preferably locked and accessible to as few individuals as possible. For sensitive personal data in public documents, secrecy is valid for a maximum of 70 years from the day the document was made public. All rights for secrecy must be tried in each separate case.

If you have any questions regarding access to information and secrecy contact the responsible person at your department, the university archivists or the legal advisors at the University Administration.

Biobanks in Medical Care Act

In 2003 the **Biobanks in Medical Care Act** (Lag om biobanker i hälso- och sjukvården m.m., also known as Biobankslagen or the Biobank Law) took effect in Sweden. The act is applicable to human biological samples taken within the healthcare system in Sweden and traceable to the humans from which they originate. The act regulates how human biological material taken within healthcare is stored, handled and used with respect to the individual's integrity. Before samples are collected for research an ***ethical approval*** is needed and can be obtained from a ***Regional Ethical Review Board*** (regional Etikprövningsnämnd, EPN).

Biobanks (sample collections)

Biobanks are defined as a "collection of biological material, stored for one or more reasons including information about this material". Sample collections at KI, whether they are kept by the researchers themselves at the department or at the centralised core facility (***KI Biobank***), are all part of the official KI Biobank and the researchers themselves do not need to report the sample collections to ***the National Board of Health and Welfare*** (Socialstyrelsen), this will be done by KI Biobank when necessary. If you plan to start a new sample collection, please contact Cecilia Björkdahl (cecilia.bjorkdahl@ki.se, 08-524 872 65) at KI Biobank for information about the local rules at KI and for internal registration.

Additional information about biobanks and biobank samples and collections can be found on the KI Biobank webpage (http://ki.se/ki/jsp/polopoly.jsp?d=9620&l=en[19]) or at the webpage of Biobanksverige (Biobank Sweden, www.biobanksverige.se[20]) where ***the National Biobank Committee*** (Nationellt Biobanksråd) supplies information about biobanks and biobank samples.

[19] KI Biobank, http://ki.se/ki/jsp/polopoly.jsp?d=9620&l=en

[20] Biobank Sweden = Biobanksverige, www.biobanksverige.se

In the future, all biobank sample collections in Sweden should be reported to ***the Swedish Biobank Registry*** (Svenska biobanksregistret, SBR) – if you have any questions please contact Cecilia Björkdahl at KI Biobank.

Informed consent

Everyone that donates a biological sample or information to research has to give an ***informed consent***. The research subject should be informed of and consent to:

- participation in the study,
- collection of biological samples (if applicable) and,
- handling of personal data (this may include copies of or information from their medical journals).

The Ethical Review Act states that the research subject should be informed about:

- the overall plan for the research,
- the aim of the research,
- the methods to be used,
- the consequences and risks that might follow from the research,
- the responsible principal entity (ansvarig huvudman),
- that participation in the study is voluntarily and,
- the right to withdraw his/her participation/consent at any time.

The Ethical Review Board needs to approve the finalized versions of the information about the research and the informed consent.

The Personal Data Act states that the research subject should be informed about:

- the organisation responsible for the personal data, including contact information for the controller (address and telephone number),
- a contact person the research subject can get in touch with,
- the purpose of processing the data,
- whom the information will be shared with (if at all), this is of extra importance if the personal data will be sent outside the EU/EEA (EEA = European Economic Area, EU/EES-området), e.g. to the USA,
- how long the personal data will be kept,
- the level of confidentiality for the information,
- and the right for the research subject to:
    - obtain a printout of one's personal details,
    - have incorrect information corrected and
    - suspend the registration.

The personal data representative (controller) is ultimately responsible that personal data is handled correctly, meaning that correct information according to the Personal Data Act is given to the subjects and that expressed consent about the registration is collected. The consent should preferably be collected in writing. If given orally, the consent should be well documented. For the record, it is also good to state why no written consent exists.

Research without informed consent

The Regional Ethical Review Board may approve research without informed consent; this might apply, for example in registry based research. The Regional Ethical Review Board may then decide if and in what way information should be provided to the research subjects. In some cases, this can be done through an ad in the daily press, for example, which gives the participant the opportunity to become informed about the research and withdraw from it, if desired.

Biological samples and/or personal data that leaves KI

If coded samples are sent away from KI for analysis, not just abroad but also to other principal entities within Sweden, a ***Material Transfer Agreement*** (MTA) should accompany the samples. If the transfer also includes personal data or analysis results (that can be traced to an individual), this too needs to be documented, either in a separate contract, a so called ***Biträdesavtal***, or included in the MTA.

The MTA usually includes instructions on handling samples and personal data, as well as security and secrecy rules and regulations that apply. It is also specified when the material (information and/or samples) should be returned or destroyed and who is responsible for paying any possible resulting damages.

If someone asks to be given access to samples or personal data from KI this should be put to trial (menprövning). If, when tried, no reason (e.g. secrecy) prevents access to the samples, a MTA should be signed. In some cases the personal data should be regulated by a secrecy contract (sekretessavtal). Secrecy contracts are usually only needed when the request for access comes from the private trade and industry.

If personal data, or coded samples including personal data, are sent outside of the EU/EEA the research subjects must have given their (informed) consent to this. The Regional Ethical Review Board does not have the mandate to approve this.

Tissue directive (*Vävnadsdirektivet*)

If the research includes transplantation of human tissues and cells in humans, the work needs to comply with the EU Tissue directive (Vävnadsdirektivet) and the corresponding Swedish legislation[21]. The Tissue directive sets standards of quality and safety for donation, procurement, testing, preservation, storage and distribution of human tissues and cells intended for human application[22].

Animal Welfare Act

An ethical approval is needed for all research that includes animals. In Sweden this includes not only live animals that are part of the experiments, but also animals that are killed solely for

[21] Tissue directive = Vävnadsdirektivet, Lag om kvalitets- och säkerhetsnormer vid hantering av mänskliga vävnader och celler (2008:286)

[22] Quality standards for human tissues and cells – summaries of EU legislation, http://europa.eu/legislation_summaries/public_health/threats_to_health/c11573_en.htm

harvesting organs for further experiments. According to ***the Swedish Board of Agriculture's***[23] webpage, laboratory research animals are *"animals used for scientific research, disease diagnosis, development and manufacturing of pharmaceutical or chemical products, some kinds of education and other similar purposes*". More detailed information can be found at the Swedish Board of Agriculture's webpage (www.sjv.se), at ***Djurförsök.info*** (http://djurförsök.info) and in the legislative act itself.

Handling of laboratory animals is regulated in the **Animal Welfare Act** (Djurskyddslagen) and permission to use laboratory animals is required from both the Swedish Board of Agriculture (Jordbruksverket) and an ***Ethical Committee on Animal Experiments*** (Djurförsöksetisk nämnd). Each animal facility at KI has their own permission from the Swedish Board of Agriculture, so KI researchers only need to apply for a research specific permission from an Ethical Committee on Animal Experiments. The application to the Ethical Committee on Animal Experiments should also be signed by **Godkänd föreståndare** (approved supervisor of the animal facility) and be reviewed by the **Veterinary department** (Veterinärenheten) at KI.

Everyone wanting to work with laboratory animals at KI needs to attend a **Laboratory Animal Science course** (Försöksdjurskurs) at KI or equivalent. Before attending the course, work in the animal facilities at KI is only allowed (for a maximum of 2 years) under supervision by someone qualified according to KI's rules. The guiding principles for laboratory animal science are **the 3 R:s**: **replacement** of living animals, **reduction** of the number of animals and **refinement** of the experimental procedures[24]. For all animal based research at KI documentation should be completed for all animals handled by the researcher and/or employee at the animal facility. This documentation should be stored for at least 3 years after the animal is killed or the experiment is finished.

For more information on laboratory animals and animal based research, please contact veta@admin.ki.se. In case of an animal emergency, please call 08-524 867 22.

[23] Information about laboratory animal science from the Swedish Board of Agriculture = Information om försöksdjur och forskning från Jordbruksverkets hemsida, http://www.sjv.se/amnesomraden/djurveterinar/djurskydd/forsoksdjur.4.b1bed211329040f5080003128.html'

[24] The 3 R:s = de tre R:en, http://intra.ki.se/service/veterinary/intro/intro2_se.html

What archiving demands are there?

Archiving Plan[25]

All public documents received or produced at KI are, as mentioned above, considered to be material owned by KI. This means that the material must stay in the archives at KI. Researchers are not allowed to decide by themselves which research materials or documents can be destroyed. This is regulated in the **Archive Act**[26] (Arkivlagen) and ***the National Archives*** (Riksarkivet), who supervise archive management throughout Sweden. The archivists at KI have therefore published an **Archiving Plan** (Arkivbildningsplan) as a guide for keeping and archiving research documentation at the departments at KI.

Extract from the Archiving Plan:
"Documents received or produced by a KI department are considered public documents and shall be preserved by law to meet: a) the right to access public documents, b) the need for information for the administration of justice and public administration and c) research needs."

It is the responsibility of the head of the department to ensure that correct archiving routines are followed and that the archives at the department are stored in approved archiving facilities.

Registering and archiving official documents

In order to facilitate the public's exercise of the right to access official documents (which can be classified as public or secret), it is important for the public to be allowed to know which documents are held by the public authorities. Consequently, official documents received and/or produced by KI must be registered. Regarding research material, such documents are e.g. permits and documents regarding external financing (e.g. contracts for research funding or granted applications for funding).

There are, however, some exceptions to this rule of registration. This means that some documents produced in the process of research must be registered and some documents can be put directly in the archive without registering. Research is consequently archived in different ways, and is divided into;

- registered documents (both preserving and non-preserving material put in the archive),
- preserving material
- non-preserving material

Documents containing secret information that is classified as secret according to The Public Access to Information and Secrecy Act must always be registered. Registration is done by each department. Details on how to register can be provided by your supervisor or the head of administration (administrativ chef, AC) at your department.

[25] Archiving plan = Arkivbildningsplan för institutionerna vid Karolinska Institutet (Dnr:4159/08-040)

[26] Archive Act = Arkivlag (1990:782)

All documents regarding research shall be kept within KI. You may not take these documents home. If doing so, you are obligated to return to KI with the documents if the public requests to see them. This obligation is irrespective of vacation or illness.

Listed below are three classifications of documents/records (handlingar):

1) Registered information (documents)

Examples of information that must be registered are:

- Documents containing secret information
- Contracts
- Agreements on cooperation, collaboration, etc.
- Applications concerning genetical analysis

2) Records of long-term value

Examples of research material that usually is not registered, but has to be preserved in a ***project file*** (projektakt) are:

- Project plan(s)
- Permissions/approvals (e.g. from the Ethical Review Board or the Swedish Data Inspection Board) and documents regarding external financing (shall be registered)
- Primary data (both from registries and original raw data)
- Minutes from larger national/international collaborations
- Internal correspondence of principal importance for valuation of methods and results
- Other unpublished reports
- Templates for questionnaires that are sent out
- Finalised list of publications

3) Records for disposal (non-current records)

As mentioned, not all documents need to be preserved forever; some can be sorted out according to the rules in KI's **Archiving Plan**. The regulations state *when* you can sort out the documents. Not all documents can be sorted out at the same time. Some documents can be sorted out directly, others must be saved until your project is completed, and other documents must be saved for a period of 10-15 years.

Some working material (if not officially archived), abstracts and manuscripts ready to be printed/accepted can be sorted out when they become outdated or obsolete, while others can be sorted out at the end of the project (e.g. imported registry data and information from biobanks or patient/medical journals).

Examples of documents that may be disposed of after 10 years (15 years if they involve clinical trials [läkemedelsprövning]) are:

- Documents like informed consents
- Laboratory notebooks
- Experimental protocols

- Quality controls
- Changes/corrections of primary data
- Photographs and printouts from laboratory equipment/measuring apparatus etc.

Information that, according to the archival regulations, is to be preserved for the future must be separated from the information that is supposed to be destroyed. Information allowed to be destroyed should be clearly marked. This is important, as you might not be available at KI to sort them out by yourself in the future. Check with your supervisor or the head of administration (administrativ chef) at your department where to put the material and how to do it.

When the destruction of secret information is warranted, it is important to destroy the secret information safely. You may not simply throw secret documents away. It is your responsibility to make sure that secret information does not get into the wrong hands. It is usually only duplicates of the secret information that can be destroyed, while the original information is preserved. The decision to classify information as secret is always determined in legislation or by legal advisors at KI. All secret information must be registered before destruction.

Electronic information

Today all archived information received or produced by KI should be preserved on paper, but this may change in the future. Some exceptions are already made for research databases due to their size (see the Archiving plan for approved formats). Upgrades or conversions that results in loss of information are forbidden, and must be thoroughly investigated with special consent from the National Archives. All contact with the National Archives must be done by/via university archivists at the University Administration.

All research databases on paper must be registered by the department. Check with your supervisor or your department administration on how to do this. The electronic databases must be reported to the ***Personal data representative*** (Personuppgiftsombudet) at KI for registration and the application form will be archived centrally at KI.

Archival repositories (Arkivlokaler)

All material that is preserved at KI must be put in rooms specially designed for protecting archives. These rooms have special climate regulation systems. Electronic media and paper materials have different needs for preservation. You are allowed to keep material that is decided to be destroyed within a period of 10 years in your own office (non-preserving materials). Material that must be preserved for a period of 10 years or longer, must be put in a room designed for protecting archives.

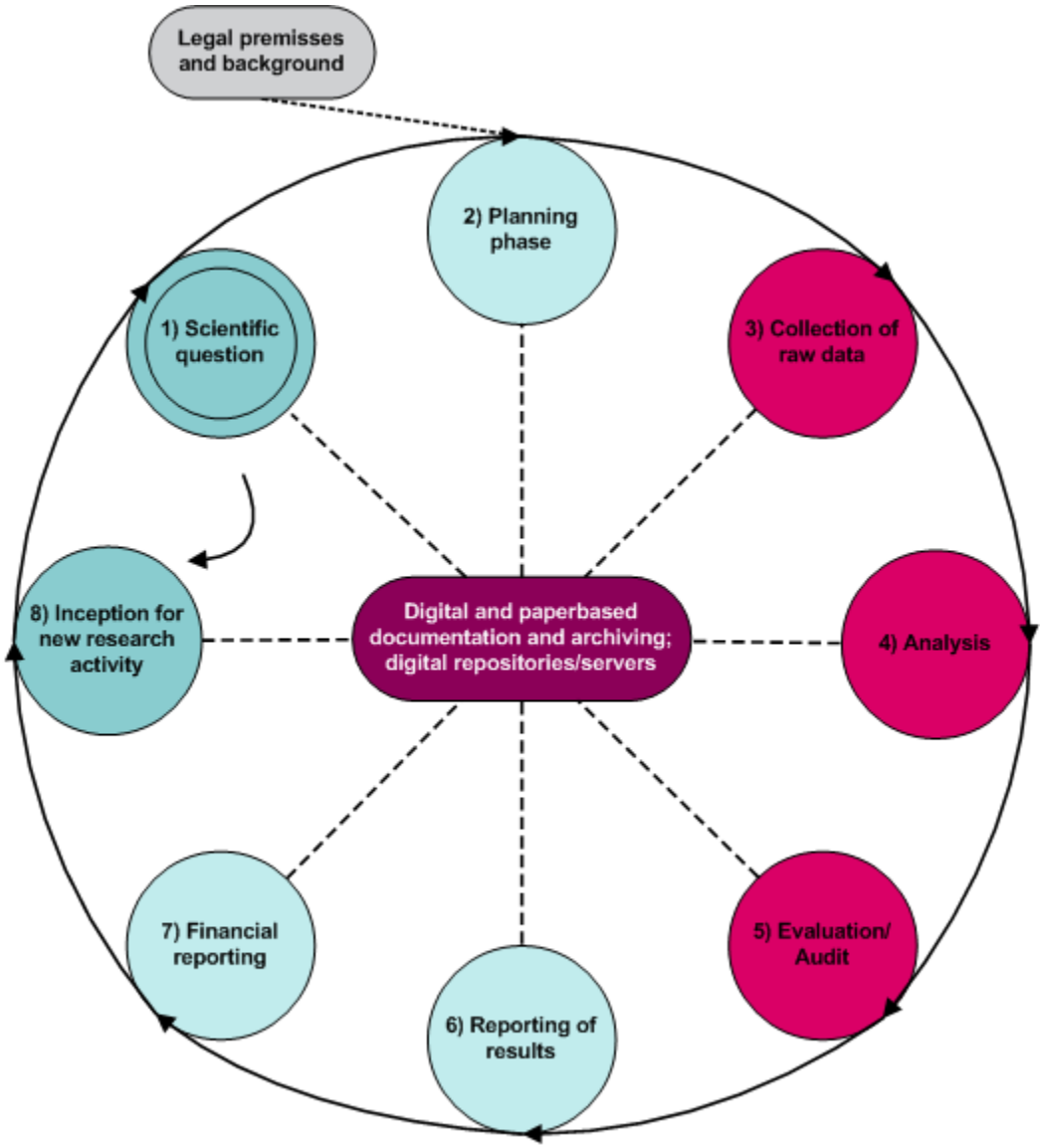
You are not allowed to put material outside of KI. Private organisations financing your research are not entitled to take over the material when your research is completed, and may not borrow material either. All material should be held within KI.

If you leave KI…

Every now and then researchers want to complete their research elsewhere, either at other universities or at private organisations. In these cases, you are not allowed to take your research material with you. If you switch universities or work for another public governmental organisation (inom staten) in Sweden you might be allowed to take the research with you. That depends on several factors and is always decided by the university archivists. If you move to another country you cannot take your research with you at all, nor if you go to any private organisation.

If you have any questions, please contact the university archivists at the University Administration.

Documentation during the research process



Adapted from original from Renata Arovelius, SLU

Definition of research data

Various organisations and universities have defined research data along the following lines and more detailed information can be found, for example, on the University of Edinburgh's

webpage[27]. Their definition of research data is: "Research data, unlike other types of information, is collected, observed, or created, for purposes of analysis to produce original research results."

They also classify research data as:

- ***Observational***: data captured in real-time, usually irreplaceable.
- ***Experimental***: data from lab equipment, often reproducible, but can be expensive.
- ***Simulation***: data generated from test models (where model and metadata are more important than output data).
- ***Derived or compiled***: data is reproducible but expensive.
- ***Reference or canonical***: a conglomeration or collection of smaller datasets, most probably published and curated.

The research data comes in many different formats – both electronic, paper and figure based.

These definitions clearly apply to the research done at KI and may serve as an additional reference when deciding what to document.

What should the documentation include?

Different parts of the research process have different demands when it comes to research documentation. Some of the more important aspects of the research that should be documented are listed below, numbered according to the above figure[28]. *At a minimum, the information in steps 1-4 should normally be included in the laboratory notebook (log book, journal).*

It is important to be able to follow the researcher's train of thought, experiments, observations, analysis, etc. leading up to his or her final results and conclusions. Usually, it is also good to document why something was NOT done, why another option was used, etc. The process(es) from the planning stage to the final report should be evident, documented in a good way and traceable if the need arises to go back to prior results and analyses – this is needed when experiments are repeated and when obtained results are reconstructed retrospectively. Any alterations or exclusions/additions made to the data need to be documented, dated, signed and justified.

It is always good to document the research continuously, with enough detail to allow outside scrutiny. Final reports should be written as soon as possible after completion of the experiment/analyses/project.

At KI all research documentation must be written in Swedish or English, and be intelligible to an outsider.

[27] http://www.ed.ac.uk/schools-departments/information-services/services/research-support/data-library/research-data-mgmt/data-mgmt/research-data-definition

[28] This figure is adapted from work on European guidelines for research documentation together with Renata Arovelius at the Swedish University of Agricultural Science (Sveriges Lantbruksuniversitetet, SLU).

1 – SCIENTIFIC QUESTION

- Based on previous results and raw data. New ideas/hypotheses.
- References to earlier research, the "thought behind" (not just what you are planning to do), motivation, validation.

2 – PLANNING PHASE

- Research and/or project plan. Background and aim(s) with the research/project.
- Description of the method(s) used.
- Financial and administrative records. Grants applied for (and received).
- Ethical applications (for humans and/or other animals).
- Preservation strategy: archiving, IT solutions and security.
- Metadata planning – which are needed/provided and why.
- Applications for access to material, e.g. registry data and/or biobank samples.
- Documentation of decisions made that affect the project.

For clinical and epidemiological research the documentation should allow scrutiny of whether the selected sample of research subjects is representative of the original population from which they have been recruited. A clear description is required of both the selection criteria and the means by which the research subjects are enrolled in the study. Management of dropouts, should they occur, should also be described at this stage.

3 – COLLECTION OF RAW DATA

The documentation requirements at this stage are discipline dependent, but should be detailed enough to enable traceability and reproduction of the research. In clinical/epidemiological projects it must be possible to clearly identify primary/raw data for individual research projects and the data must be filed securely (inaccessible to unauthorised persons). Secrecy of personal data must be respected and, where possible, coded lists should be kept.

Should include information about:

- Date
- Person responsible for work, notation etc… ("signatures")
- Protocols, instructions, SOP:s
  - Which methods are used and why?
  - Validation
- Starting material
  - Kind or type of sample: biopsy, homogenate, cell lysate, purified protein or DNA, etc.
  - Species: human, animal, cells, etc.
- Materials
  - Chemicals, buffers, solutions, antibodies, kits, etc.
  - When relevant include lot and batch number
  - Quality control

        - Who prepared a solution/buffer/antibody/etc. and when
        - If the solution/buffer/antibody/etc. can be reused, how many times it has been used
    - Laboratory apparatus
        - Maintenance, calibrations and validation(s), quality control
        - What machines are used and why
        - Default settings
        - Whether the apparatus/machine (e.g. FACS, plate reader, Mass Spec, microscope) uses any software programmes for calculations, etc.
    - Variable description(s)
    - Coding
    - Metadata
    - Data sources
        - Existing
        - New
        - Experimental
    - Measurements, data collection
        - Metadata
        - Validation
        - Quality control
        - Standards, min and max values/limits
        - Programmes and/or methods used
    - Calculations
    - Etc…

4 – ANALYSIS

Raw **and** processed data:

- Selection criteria
- Synthesis
- Calculations
- Statistics – methods and software programmes used. Documented logs or similar for selection and analysis.
- Interpretation
- If raw data is altered in any way this should be clearly described and justified.
- Etc…

It should be possible to assess experimental variation within and between experiments, reproducibility and possible methodological errors.

A lot of the material/documents during this step are so called working material and are usually not considered public documents. Some working material has to be stored because it is necessary in order to understand other material/texts/etc. or for a specific context. If it is stored and archived, it will become a public document.

5 – EVALUATION/AUDIT

Peer review:
- Seminars
- Conferences
- Scientific discourse
- Publishing

6 – REPORTING OF RESULTS

- Sub-reports
- Final reports
- Media messages
- Popular science
- Papers

7 – FINANCIAL REPORTING

- Grants and other funding received
- Statement of accounts, financial report

8 – INCEPTION FOR NEW RESEARCH ACTIVITY

- Reuse of raw data
- Infrastructure

Advice and help on how to document your research

As the level of detail usually determines the ease with which an experiment or analysis can be reproduced, do a "fire drill" and test if someone unfamiliar with the project/experiment can reproduce it from the information included in the available documentation. Below follows a selection of advice for different kinds of research.

If raw/original data is changed or modified in any way this needs to be documented, traceable and motivated/explained. It is also good to always keep an "untouched" version of your raw/original data that you can go back to if necessary, e.g. if you need to redo the analysis.

Keeping track of antibodies, chemicals, pre-made solutions, etc. is easier if a logbook is kept. The information in the logbook should include:
- Manufacturer/provider
- Date of delivery
- Amount: volume, concentration
- Results from validation and/or optimisation tests
- Dilution (also when, what kind of solution and by whom)
- Reused: number of times
- And anything else you deem necessary…

Put together specific "manuscript/paper folders" where all the information (or copies thereof) that goes into a manuscript is included, e.g. references to original and adapted images (this is especially important if one specific image is used to represent the whole experiment, you should be able to find ALL images that form the basis for the manuscript).

Standard protocols, SOPs, etc. can be reused or just referred to in the notebook – as long as any deviations or changes from it are well documented.

If other people handle your samples, e.g. personnel at the animal department or hospital staff, make sure that you keep track of the whole "chain of custody". This can be done, for example, by special coding designed to give information about who has handled the sample and if it has been treated in any way.

Differences between research fields

The best way to document research will vary depending on the research field and kind of research performed.

Experimental/pre-clinical research

For experimental/pre-clinical research, the details will vary a lot depending on the chosen topic/subject and the researchers themselves usually know best what information should be included. If you are unsure as a PhD-student on what to include, make sure to ask your supervisor for help and details. Most experimental/pre-clinical research can readily be documented using a (paper or electronic) laboratory notebook.

An example of how instructions and details for the documentation can be expressed is provided in Appendix A. This document has been put together by Dr Camilla Svensson at the Department of Physiology and Pharmacology (FyFa) and is used in her group. It provides a very good example on how to document your research.

Epidemiological research

Epidemiological researchers, especially when the research is questionnaire and/or registry based, have a harder time utilising a laboratory notebook since most of the data is handled in an electronic form. This has prompted the Data Management Group (DMG) at the Department of Medical Epidemiology and Biostatistics (MEB) – a department doing mainly this kind of research – to put together guidelines on how to **"Take good care of your data"**[29] and **"Documentation and archiving of computer media files in research projects"**[30].

[29] MEB version of Take good care of your data by Svend Juul, with contributions from Jens Lauritsen and Annette Jørgensen, adjustments to the MEB-environment by Åsa Klint and Anna Johansson, May 2006, MEB

[30] Guidelines for Documentation and Archiving of Computer Media Files in Research Projects", version 2, 14 November 2006, MEB

In short, these guidelines state that:

- The design of the data collection is important for the value of the data collected.
- It should always be possible to trace each piece of information back to the original document
  - The origin of the data.
  - ID (case identifier) included in the original documents and data set.
  - All corrections must be documented and explained.
  - All modifications of the data set must be documented by command files.
  - A command file must document each analysis.
- Organise and describe a folder structure that should include folders for:
  - Data – data files from the program(s) used
  - Documents – documentation files such as the research plan, ethical application(s), logbook and codebook files, manuscript versions, etc.
  - Log – log files from data preparation and analyses
  - Output – output files from data preparation and analyses
  - Programs – program files from data preparation and analyses
- Decide on a system for naming data and command files.
- A **code book** is compiled were the name, meaning and coding of each variable is described and may serve as a link between a questionnaire and data entered in a computer, for example.
- Keep track of study participants.
- Take precautions when entering data – such as setting up a data entry form resembling the questionnaire, define valid values before entering the data and enter data twice – to prevent errors. Check the data for errors.
- A **log book** should be kept were modification of or changes to the data are documented, this may include error corrections, data cleaning, etc.
- Do not modify the original data, make working copies instead.
- Make sure that the right data set is used for the analysis.
- Be sure to back-up your data regularly and test if you can retrieve it after e.g. a (simulated) computer crash.
- Archive the necessary parts when done, either electronically or on paper according to the regulations.
- Encrypt the data when necessary.

For further information contact the DMG via Rozita Broumandi (rozita.broumandi@ki.se) or Sandra Eloranta (sandra.eloranta@ki.se).

Clinical research

It may also be hard to keep a traditional paper laboratory notebook in clinical research. But the research still has to be documented. If most of the notations and test results are done directly in the medical journal, additional information is needed on what kind of information is documented in the medical journal and by whom. For these “logs” a paper laboratory notebook may suffice. Personal data may also be documented outside the medical journal, e.g. in data registries and in paper laboratory notebooks, as long as these other “formats” are handled according to the usual laws (the Personal Data and Public Access to Information and Secrecy Acts, for example).

If larger data sets are analysed the clinical researcher may benefit from the same guidelines as for questionnaire and registry based research (see above) since it is important to be able to understand how data was handled and the results were obtained.

Electronic help with documentation

Most researchers today depend on word and image processing software, not to mention the various statistical programmes and those necessary for the different analysis apparatus found within the laboratory environment. With an increasing amount of the research data being kept and collected in electronic form it is not surprising that electronic tools for handling data are becoming more and more common.

For more detailed information on specific handling of digital data, please see Appendix B.

Laboratory Information Management System (LIMS)

Researchers are taking advantage of so called Information Management Systems, especially of the laboratory kind (Laboratory Information Management System, LIMS). LIMS are particularly good for structured data and large data sets like databases and/or patient information.

Today a LIMS has been purchased jointly by KI and the Stockholm City Council (Stockholms Läns Landsting, SLL) and can be used under the name of SCARAB[31] by researchers at KI. To date, the majority of the LIMS users at KI work with patients and patient material.

For more information about LIMS and SCARAB, please go to the SCARAB webpage (http://ki.se/scarab).

Electronic Laboratory Notebook (ELN)

For experimental/pre-clinical researchers, moving from a paper laboratory notebook to an electronic version (Electronic Laboratory Notebook, ELN) may help to improve the quality of the research documentation since both traceability and the possibility to search the records are greatly enhanced.

KI will implement a central ELN-system in 2010. If you have any questions and/or are interested in moving from paper laboratory notebooks to an ELN please contact Cecilia Björkdahl at 08-524 872 65 or cecilia.bjorkdahl@ki.se.

[31] SCARAB = Sample Collection, Administration, Research and Biobanking. The name chosen for the LIMS implementation at KI and SLL and the group working with it, see http://ki.se/scarab for more information.

Acknowledgements

The author wish to thank the following people for their valuable input and expert knowledge that helped in finishing this handbook:

At KI, Kerstin Bogren (personal data representative), Heléne Lindström (university archivist) and Solveig Tjäder (head veterinarian) have all helped with ensuring the accuracy of the information within their respective fields. Also various researchers and the working group for research documentation (appointed by the Board of Research in December 2008) have been very forthcoming with their ideas and examples on how to document research. Special thanks also to group leader Camilla Svensson at FyFa for sharing her "Notebook rules and guidelines". The author also wishes to thank Kate Svensson for her proofreading and helpful suggestions of the handbook draft.

Outside of KI, Renata Arovelius (head archivist at the Swedish University of Agricultural Sciences in Uppsala) has been an extremely valuable source of information. She has also introduced collaboration between KI and the Committee on Scientific and Research Data – a committee working on a European handbook for research documentation where KI's work will serve as an example.

Contact information

If you have any questions regarding this handbook or suggestions for additions and/or revisions, please contact Cecilia Björkdahl at 08-524 872 65 or cecilia.bjorkdahl@ki.se.

Appendix A – CIS Laboratory Protocol, Nov 29 2008

NOTEBOOK RULES AND GUIDELINES IN CIS LABORATORY

The laboratory notebook is a permanent, documented, and primary record of laboratory observations. Therefore, your notebook is a journal with pages that are numbered in advance and should never be torn out. All documentation in the lab book has to be in English. Write your name, group (Molecular Pain Research), Department (Physiology and Pharmacology) and start date on the cover. All notebook entries must be in ink and every page clearly dated. Provide the full date whenever you make an entry. Avoid 02/2/02 and 2/2/02 type dates, especially as our laboratory has a mix of nationalities and dates are written differently. Opt for a format that leaves nothing to guesswork: *for example 2 Feb 2002; Feb 2, 2002;*

In view of the fact that a notebook is a **primary** record, data are not copied into it from other sources (such as a lab partner's notebook in a joint experiment) without clear acknowledgment of the source. Observations are never collected on note pads, filter paper, or other temporary paper for later transfer into a notebook. If you are caught using the "scrap of paper" technique, your supervisor will be very upset! It is important to develop a standard approach to using a notebook routinely as the primary receptacle of observations.

The notebook is written in chronological order, i.e. an experiment that stretches over several days or weeks will be interrupted by other experimental notes. When coming back to an experiment that has already been introduced simply rewrite the Title of the project and add “continuation from p. xxxx”.

Each experiment should include the following:

1. TITLE
Every experiment should have a title and it should be descriptive. An example would be *"Assessment of Gadd45b protein expression in IL-1 and TNF stimulated astrocyte cultures".*

2. BACKGROUND AND AIM:
This section should include any information that is pertinent to the execution of the experiment or to the interpretation of the results. The background section of your notebook is not like the introduction to a paper. Include anything that will be helpful in carrying out the experiment and deciphering the experiment at a later date. For the most part, notebooks are not written for today but for the future.

For example, if it is a behavioral experiment based on a finding in astrocyte cell cultures, include what the previous finding was (yours or cite a paper or even a discussion with someone), an example would be: ”In my previous work I found that TNF stimulation of spinal primary astrocytes increases Gadd45b protein expression (See page 450855 note book 1). As TNF is elevated in CSF 4 hours after injection of carrageenan to rat paw (Andersson AA et al J. Neuroscience 2008), we will examine if Gadd45b protein expression is increased also in the spinal cord at time points when TNF concentrations are elevated). Camilla talked to Dr Tony Yaksh at UCSD (tyaksh@ucsd.edu) at the SFN meeting 2008 and he suggested that we assess Gadd45b protein expression by ELISA rather than by WB, as he has found that more reliable. However, my Gadd45b WB worked well in astrocyte cell cultures and therefore, the aim with this study is to run both ELISA and WB on spinal cord tissue in parallel, in order to determine which

system is best for our type of work. I will use extracts from Gadd45b over-expressing C6 cells as a positive control."

3. MATERIALS:
This section should include any materials, i.e. solutions or equipment that will be needed. Composition of all buffers should be included unless they are standard or are included in a kit (and then the kit name, vendor, and catalog number are given). Include all calculations made in preparing solutions. Biological and chemical reagents should be identified as to **names, sources (i.e. vendor) and catalog numbers,** and when appropriate, lot/batch numbers. For animals, include as much info as possible, i.e. mouse strains, source, animal facility, ID, weight, overall condition (animal work will be covered in more details in a separate protocol). Whenever water is used, specify de-ionized, distilled, tap, cold, hot, sterile, etc.

4. PROCEDURE:
Write down **exactly what you are going to do** before you do it, and make sure you understand each step before you do it! You may start by including the lab master protocol (or referring to it) and then your own modified one. You should include everything you are planning to do, including volumes and amounts. It is OK to paste in a print out of your protocol for the experiment as long as the correct dilutions and calculations are on there, or you left blanks to fill. However, writing a procedure out helps you to remember and to understand what it is about. It will also help you to identify steps that may be unclear or that need special attention.

Some procedures can be several pages long and the protocols may include more information than is necessary for your experiment, for example the manual from the manufacturer. In those cases you should write a short form protocol. If an experiment is a repeat of an earlier experiment, you do not have to write down each step again, instead refer to the earlier experiment/protocol by page number. For experiments that you are doing on a regular basis, use good laboratory practice and have a separate binder for such methods, so that you have easy access to those protocols while in the lab, and do not have to carry around multiple note books.

If you make **any** changes to the protocol, note the changes and why. Flow charts are sometimes helpful for experiments that have many parts. Tables are also useful if an experiment includes a set of reactions with multiple variables. It is good practice to check off steps as they are completed, or reagents as they are added, to prevent you from losing you place or for forgetting to add something.

5. EXPERIMENT:
Take **notes during the experiment**, note anything that deviates from your PROCEDURE plan. Note observations and calculations along the way.

Examples: cell number counted and calculation for volume of cell suspension to add to each well to get X number of cells/well… or "Mouse 3D did not support on back left leg, difficult to test"… "I suspect I mixed up sample 6 and 9"…

6. RESULTS: This section should include **all raw data**, including gel photographs, printouts, cell counts, IHC pictures or comments and references to glass slides, behavioral measurements etc. This section should also include your analyzed data, for example, transfection efficiencies, calculations of protein/gene expression changes, enzyme activities and behavioral assessments and weight changes in animals. Any electronic files generated as experimental results, for

example WB image files or sequence chromatograms should be filed with the general laboratory data files on your computer and the lab data file on the FyFa server.

7. CONCLUSIONS/SUMMARY:
This is one of the most important sections. You should summarize all of your results, even if they were stated elsewhere, and state any conclusions you can make. If the experiment didn't work, what went wrong and what will you do the next time to try to trouble shoot?

Sign off by you and your supervisor

Notebooks will be read and signed by you and your supervisor on a regular basis (should be signed for the previous month by the 5th. When notebooks are examined, we will look for the following points:

- **Overall readability.** Therefore - write legibly! If you ever use your laboratory notebook to reconstruct experimental details for use in a manuscript, poster, or seminar presentation, you will be miserable if you haven't been reasonably neat. Your notebook does not have to be a work of art, but it should be easily readable by *another scientist.*

- **Prelab write-up** that shows that you were prepared for lab *before* beginning the experiment.

- That calculations and graphs are **correct.**

- **Enough explanatory information** so that anyone else in the lab could, *from your notebook alone*, enter the lab and repeat your work.

- That you have **summarized your data and included a discussion**

Other notebook rules and guidelines

Typing errors
No entry is ever erased or obliterated by pen or "white out". Changes are made by drawing a thin single line through an entry in such a way that it can still be read and it is clear where the new entry is. If it is a primary piece of data recording that is changed, a brief explanation of the change should be entered (e.g. "balance drifted" or "reading error"). No explanation is necessary if a calculation or discussion is changed; if a whole section is deleted simply remove it by drawing a "x" across it.

Do not forget units!
A laboratory notebook should be legible, and data in it should be readily accessible, clearly labeled with units, and grouped in a logical way.

References, collaborators etc
Because some information might come via e-mailed suggestions (from your mentor or colleagues) or from online sources (PDFs or web sites), you might paste "miniaturized" versions of relevant passages directly into your notebook (rather than spending the time to transcribe).

Record names of people providing assistance with data collection, techniques, statistical advice, equipment loans, stipend support, supplies funds. Write this information down immediately so that you remember to include it in your future "acknowledgements" sections. Keep notes about phone conversations.

Record reagent details
Repeating this one because it is so easy to forget. It is extremely important to know where the reagents came from and the specifics of them. The following details should be included: name of reagent, vendor and product number (when appropriate, also chemical structure, purity grade, pH, lot/batch number, date of mixing/production, expiration date, etc.). Whenever water is used, specify de-ionized, distilled, tap, cold, hot, sterile, etc.

This is your lab diary!
Make daily entries, even if to say just "checked astrocytes, looks good."

THE NOTEBOOK CAN NEVER LEAVE THE KAROLINSKA INSTITUTET
Typically, the notebook should stay in the laboratory and never be brought outside the lab hallway. However, it is sometimes necessary to bring the notebook to for example the vivarium, the pharmacology house or another department if the experiment is performed there. However, NEVER bring your notebook home or elsewhere outside the KI campus. If you want to bring information from your notebook, use the copy machine or take photos.

Last thing, do not leave your notebook on your desk or in the lab over night. Place it in a drawer in your desk.

Good luck with your work!!

Read and understood by: _____ date:_____

Printed name: _____

Appendix B – handling of digital data

Handling of digital data

Computers are developing by far faster than any legislation, hence it can be hard to interpret and apply the regulations. The current best practice for some common cases is explained here.

Image data

An image should in a research setting be considered a measurement, not a pretty picture to be put in a frame. This requires an understanding of what a digital image really *is*.

The primary content is a grid of color intensities (*pixels*). These values are normally integers in range 0-255 (8-bit), 0-4095 (12-bit) or higher. The range decides how precise intensities can be stored and is referred to as *bit depth*. Consumer hardware and software can usually only handle 8-bit images.

Microscope images also contain *metadata* (”data about data”). This can be anything depending on the setup, but typically *resolution* (number of pixels per micrometer), when the image was taken, on which microscope, temperature and so on. Not all file formats can store the same metadata. TIFF-files tend to contain a lot of metadata, but badly organized. Some file formats such as GIF and BMP does not store any metadata. Conversion from one file format to another usually means loss of metadata.

Images require a lot of disk space. For practical reasons, various *compression* methods are applied to reduce usage. These can either be *lossy* or *non-lossy*. The latter will remove some of the image content to make the images smaller than what is possible with lossy methods. The content is usually high-frequency noise but the computer does not make a distinction and it is normally up to the user to find a suitable trade-off between quality and disk usage. Each file format has a specific compression method (some none). JPEG is the most common lossy compressed format while GIF and PNG are the most common lossless formats.

The research traceability demand combined with the above leads to these guidelines:

- All handling of the image files should preserve the content. If a destructive change is made, the original should be kept.
- Resaving a file with lower bit depth is questionable.
- Lossy compression has not been accepted by the community, but may be required for practical purposes. Non-lossy compression is never a problem.
- File formats without metadata support cannot be used for research unless the original file is kept. This is an incomplete list of formats that cannot be used: BMP, PNG, GIF, JPEG, SVG, PS, PDF, XPM, PCX, AVI, QuickTime, MKV, and OGG. As a rule, if the file format is for non-researchers, it cannot be used. These formats can be used if the user manually exports the metadata and keeps it with the images, but this is complicated and not recommended.

- TIFF can sometimes be used to store images if care is taken. Due to the lack of metadata schema standardization, it is safer to use the native formats of the software used to capture the images.
- The archiving plan mandates that it should be possible to open files many years after creation. Many files cannot easily be opened today because the corresponding software doesn't work anymore, hence uncommon file formats should be avoided. Use open standard formats whenever possible. As of writing (2009), there is however no suitable open formats. Candidate formats are OME-XML, OME-TIFF and OST. If the file can be opened with Bio-formats [http://www.loci.wisc.edu/ome/formats.html] then it should be possible to open it in the future.
- It is theoretically possible to keep data as the result of an invertible operation (contrast/brightness correction) if the settings are documented but this is not a good idea since it might be hard to undo the operation many years later.
- The original data has to be kept even if the operation is almost-invertible, such as deconvolution.

Different image facilities have different rules. Some require that all data is saved as-is and is never modified, others are sloppier. Consensus is yet to be reached and your facility should have local recommendations.

Links to standards:

TIFF 6 (large part is unofficial): http://partners.adobe.com/public/developer/en/tiff/TIFF6.pdf

GIF: http://www.w3.org/Graphics/GIF/spec-gif89a.txt

OME-XML and OME-TIFF: http://www.ome-xml.org

OST: http://www.endrov.net

PNG: http://www.w3.org/TR/2003/REC-PNG-20031110/

JPEG: http://www.jpeg.org/

Other formats are not proper standards; they have been created by single vendors and then evolved.

Source Code

Source code (e.g. Java code, Perl scripts) is difficult to document with the same methods as other documents; the code is written over a long time and changes has to be continuously documented.

Changes are always written in many locations, it is infeasible to print the entire document again and again. Sometimes several people are working on the same code and it has to be clear who did what.

The industry solves these problems with Version Control. It is a specialized way of storing files that keeps track of changes. Changes are sent back and forth between all users who are reading and writing the files. The systems can either be centralized (there is a main copy, normally on a server) or distributed (all copies are equals).

There are several open source systems that can be employed, as of writing:

- CVS (http://www.cvs.com/) – one of the oldest centralized system, still works well and has good support
- Subversion/SVN (http://subversion.tigris.org/) – small incremental improvement to CVS
- GIT (http://git-scm.com/) – quickly growing in popularity and support, one of the most advanced systems. Originates from the Linux kernel developers. Distributed.
- Mercurial (http://mercurial.selenic.com) – Early distributed version control system
- Bazaar (http://bazaar-vcs.org/), Darcs (http://darcs.net/) – Early distributed version control systems, less common

Each CVS-system requires its own special client software. The major code editors (IDEs) such as Eclipse (http://www.eclipse.org/) have it built-in and are hence the easiest to use.

There are some clashes with medical regulation. One problem is that all systems allow changes to be permanently removed from the history. This turns out to be a requirement to manage large projects, especially with many external contributors. Another problem is that there is no signing of changes as it's not needed in computer science. These issues are alleviated by making backups, which should be done in either case.

Other binary data

Most software is developed for a particular purpose and is discontinued once the need disappears. This affects research greatly; some programs are published by researchers to never be updated again. Software that is not updated will sooner or later stop working – hardware changes, operating systems change, hidden bugs suddenly becomes serious etc.

To avoid a situation where the data exists but cannot be opened it must be stored in commonly accepted file formats. Whenever possible, use a format that is an open standard. Most programs will support it, and in worst case new software can be written given the specification.

Files that only can be read/written by a specific program is usually not documented enough to allow new programs to be written and companies usually do not want to give out the specification. It is a way of avoiding competition by making it impossible to change software. An example of this is the Microsoft word (.doc) format which could not be opened in other programs until the competitors managed to guess the file format structure. The specifications were finally released in 2008.

Physical digital storage of data

Data can be lost in two ways: either it becomes inaccessible or it becomes corrupted. A disk crash means data becomes inaccessible. If a bit (a 0 or 1) is changed due to cosmic background radiation then it is corrupted. Corruptions are the worst as they tend to be silent and they are hard to fix. If one copy of the data is corrupted, at least two other copies are needed to find what the correct value is. Example: If the copies are 0,1,1 then the correct value is 1. If there are only two copies, 0,1, then you cannot tell. Sometimes you can tell with a single copy; your program refuses to read a file or there is a sudden typo in the text. Luckily, disk crashes are more common than corruptions, so a single copy will rescue most situations.

Because corruptions are silent, they are likely to be saved to several backups before they are

found. A related problem is buggy software that causes data to be lost or wrong when files are resaved. The solution is to keep a history of backups e.g. once every week, 4 weeks back, and then hope the error is found before the backup is erased. To maximize the chance, do not use even time spacing; rather something like this: 1 week ago, 1 month ago, 6 months ago and 12 months ago.

Backups can be stored on disks, on optical media, on magnetic tapes or solid state memories. It is important to know the characteristics of the device you use. For normal data volumes and labs, hard drives and DVDs are most accessible.

- Hard drives are expected to work for 20 000 hours (many different numbers can be found). This is just a bit more than three years of continuous use; hence hard disk crashes are not uncommon by far. Broken hard drives can sometimes be restored depending on what is damaged but it is costly and many work hours will be lost. All hardware follows the *bathtub curve* of failure rate, meaning that many failures are early on (bad batch from the factory) or late due to wear-out. Manufacturers are expected to exercise their hard drives before use (*burn-in*) to avoid early faults; you can do extended testing by intensive reading and writing all over the hard drive (e.g. hdparm, scandisk, format with explicit checking).

- CDs and DVDs allow easy backup histories but have very limited life-span, in worst case only a few years. Make sure to read up on the type you are using and use quality brands. Avoid scratches by storing the discs properly.

Special software exists for backups. RAIDs are a useful complement. By connecting several hard drives, the computer can automatically make sure that the data can be recreated if a given number of disks are lost. RAID-5 is the most common version and is cost-efficient. Unfortunately, RAIDs do not protect against ”manual data corruption” such as accidentally deleting all files. Reliable RAID requires dedicated RAID-controller hardware with backup power; watch out for cheap software-hardware hybrids.

Redundancy is the key to reliability. So far only data copies have been mentioned but where do you store them? Keeping all the copies in your office is a bad idea in case of a fire. The media can be stolen e.g. data on a laptop. At least one copy of the data should be stored in a room which few people have access to. Using media from a single vendor or batch is a risk as they might share faults. In case of RAID, all hard drives can be lost at the same time due to lightning or the controller itself can break. In summary, you should ensure that failures are *uncorrelated events* by eliminating common factors.

The best backup routines depends on the usage pattern of each lab and how important the data is (there is no safe storage, only safer and more expensive storage). The software and hardware has to fit with the existing infrastructure. Many trade-offs between safety, performance and cost has not be discussed such as incremental backups (see e.g. RSYNC http://www.samba.org/rsync/).

Consult your IT-unit for detailed recommendations.