

EVALUATION OF DDA SUNDHED 2005-2007

The aims of this evaluation of DDA Sundhed, as prescribed in the commission letter, were to appraise

- volume and quality of the activities during the years 2005-2007, both in general and with regard to specified tasks;
- the appropriateness of the organizational solution;
- strengths and weaknesses of the efforts in relation to the needs of health science research;
- the level of activities and the scientific yield in relation to the economic resources available;
- the desirability of a continuation of the DDA Sundhed initiative.

The evaluation group has consisted of

Professor Olof Nyren (chairman), Dept of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden

Professor Mika Gissler, National Research and Development Centre for Welfare and Health, Information Division Helsinki, Finland / Nordic School of Public Health, Gothenburg, Sweden

Professor Berit Lillienthal Heitmann, Research Unit for Dietary Studies, Institute of Preventive Medicine, Copenhagen, Denmark.

The group has reviewed the 2001 report from the Committee on Archiving of Health Science Data (Udvalget Vedrørende Arkivering af Sundhedsvidenskabelige Data); the Appropriation Letter of March 15, 2005 from the Ministry of Science, Technology and Innovation; the DDA Sundhed's Plan for 2005-2008, the Annual Reports of DDA Sundhed for 2005 and 2006; and DDA Sundhed's self-evaluation "Produktion, aktiviteter og målopfyldelse 2005-2007". It also had access to the external evaluation from May 2001 of DDA Sundhed's predecessor, ERAS; as well as to the results of a survey from 2007 among documented users of DDA Sundhed's services.

SUMMARY OF THE CONCLUSIONS

In the opinion of this evaluation group

- the need for archiving procedures that result in system-independent and reasonably uniform data, transparent to external reviewers (including standardized formats for description of the research and documentation of the variables), is indisputable, mainly for the purpose of insuring against scientific misconduct (and – with respect to datasets with

personal identification – to meet the requirements from the Danish Data Protection Agency);

- the means/procedures proposed for meeting this need are sound and will not be further commented. DDA Sundhed is working on the forefront and is, among other things, developing and implementing a new documentation standard (DDI 3.0);
- the establishment of a central facility that promotes good archiving practice through influencing intra-profession opinion and offering counselling, assistance, funding and full archiving services is far-sighted and highly commendable;
- DDA Sundhed's general activity level in the period that can be assessed has been more than satisfactory considering the staff number and the allotted economic resources. Given the expected starting stretch, both for establishing the organization and gaining acceptance in the scientific community, the accomplishments so far are truly impressive;
- the efforts aimed at increasing medical researchers' awareness of the need for good archiving practices and the existence of DDA Sundhed are satisfactory, but the task is formidable. The focus of the efforts has been on visibility of DDA Sundhed and its values; the means have been a website, a newsletter, site visits, presentations in connection with various kinds of scientific gatherings, contributions to undergraduate courses and research training courses, and communication via a network of interested researchers. These activities are thought to stimulate voluntary changes for the better and to foster collaboration between DDA Sundhed and research groups. Furthermore, in the future, a certain element of compulsion through standard clauses in grants from an increasing number of leading research funding bodies is expected to result in more contacts. According to these clauses, the financial support is conditional on the researchers' compliance with the DDA Sundhed archiving scheme. The evaluation group acknowledges that the effects of the measures taken are not expected to be observable in the short term (for instance, projects funded under the new contracts will typically not be ready for archiving until after more than the 2-3 years presently under evaluation), but there are some disquieting indications of a special affinity to projects near the social sciences (which was originally the core scientific area for DDA) while clinical research so far seems to be correspondingly underrepresented. This bias is to some extent understandable since the benefits of re-using data seem to be particularly evident after surveys with collection of exposure information in broad population samples, but the availability of well-characterized consecutive patient materials would likely have a similar attraction in clinical outcome research. While the clauses requiring archiving at DDA Sundhed are found in most grants, regardless of field of medical research, the interview with Steen Ousager revealed that the ambition is presently somewhat higher as regards large-scale epidemiological and public health research, the datasets of which are perceived as typically having a greater potential for re-use. The need for assistance in archiving matters is felt to be equally pressing in all types of medical research. It appears that there is a need for intensified efforts to get the clinical research community on board. But even among researchers who were particularly expected to be aware of the DDA Sundhed initiative, the 2007 user survey unveiled a limited appreciation of the possibilities involved. There is no evidence-based information about which are the optimal channels for reaching out to the underrepresented groups, but it appears that the ones presently utilized are sensible in principle. In medical research, there is no real tradition of data sharing, but experiences from the social sciences justify considerable optimism; however,

this shift in tradition inevitably takes time. It may seem well-advised to try to bring about greater involvement among “opinion-leaders” and other key players within the underrepresented scientific areas; for instance, DDA Sundhed might consider recruiting a professional consultant (“lægefaglig konsulent”) with special responsibility for clinical research. Another idea perhaps worth considering is to form a second health science network with a “centre of gravity” in clinical research;

- the idea of promoting re-usage of data collected by other researchers for purposes not originally intended is very appealing and basically sound, but the indiscriminate promotion is limited by ethical concerns¹, ownership issues, lack of covariates needed to satisfactorily answer the new research question, and the need for a “fingertip sense” of the strengths and limitations that only the original investigators might have and which may be difficult to convey in condensed descriptions. This might be one explanation why no more than two datasets have been ordered and delivered in the almost three years under evaluation. The very small number of deliveries is much below the expectation in DDA Sundhed’s plan for 2005-2008 and could be seen as a failure. Upon direct contacts with selected researchers, it was revealed that data-sharing and re-use of data did, in fact, occur quite often, but not mediated by DDA Sundhed. Rather, the contacts were made directly between interested external researchers and the original investigators (who were presumably involved also in the new research). It is still too early to be alarmed. As pointed out above, the shift in tradition is likely to take decades rather than years. However, the development in the next 5-6 years should be carefully monitored;
- full reprocessing (“oparbejdning”) – the purpose of which is to create a fully documented and cleaned dataset ready for immediate re-use – has consumed approximately 2/3 of the study-related person-time accounted for in the self-evaluation (38% of all person-time). Particularly, internal full reprocessing not only consumes resources at DDA Sundhed, but may also be demanding for the researchers who originally collected the data. In the experience of DDA Sundhed, however, the contributions from the original investigators were often limited: in the case of small to medium-sized datasets, it was typically easier and more cost-effective for DDA Sundhed to take over the reprocessing internally, while large and complex datasets were best reprocessed externally, by the original researchers themselves. Nonetheless, external full reprocessing was labour-intensive also for the DDA Sundhed staff. This was surprising to DDA Sundhed and the cost per unit exceeded that planned for by more than 100%. The evaluation group feels less surprised and interprets the observed phenomenon as an expression of the need for substantial involvement by both parties. The evaluation groups takes itself the liberty of proposing a slight shift in the priorities according to the following scheme:
 - More resources to “primary prevention”, i.e., even heavier involvement in research training courses (why not an obligatory course in “good data management including archiving practices” at all medical/health sciences faculties in Denmark?) and stronger pressure on relevant faculties and single departments to develop and implement archiving policies. This has laudably been a prominent item on DDA Sundhed’s agenda, and more such activities are in the pipeline. Since good archiving practices should ideally be implemented already in the

¹ The ethics of a study is based on the participants’ understanding of the purpose of the investigation but also on the perceived balance between encroachment/inconvenience/pain/physical risk and the gains of the research; the original informed consent may no longer be valid and the risk-benefit equation may be quite different from the original one.

planning phase of a project, considerations to documentation and archiving should be made throughout the research process, and failure to do so will result in extra work after conclusion of the research, even more pro-active work will likely liberate person-time that is presently devoted to post-hoc documentation in the end.

- Since reprocessing of large and complex datasets will inevitably be resource-consuming at DDA Sundhed as well as in the research groups that originally produced the data – despite attempts at “primary prevention” – it is important that the selection of the most suitable datasets is optimal. Discriminate reprocessing only of the most promising ones might prevent overtaxing of available human resources and a threatening imbalance between input and output with entailing inertia and long waiting-lists – a quite realistic risk if the input of new projects continues to grow while the staff number remains on the same level. In the view of the evaluation group, the selection is best done by the medical research profession. The time devoted by the non-medical academic staff to identification and localization of suitable research projects (almost 20% of the study-related person-time accounted for in the self-evaluation, 10% of all person-time) cannot and should not be totally eliminated, but some of this time could probably be used more efficiently for other tasks (see below). Presently, the Archiving Council (“Arkiveringsråd”), the professional consultant, and to some extent the health science network are involved in the discussions, but it appears that a more formalized process might be beneficial. In analogy with suggestions in the foregoing, it might be worth considering two or more selection subcommittees in order to insure that underrepresented research areas (e.g., clinical research) will be covered. Tentatively, the members of these subcommittees should cover different sub-areas to encompass most of the relevant research activities in Denmark. The delegates might themselves come up with suggestions for suitable projects (from the list of projects turned in to DDA Sundhed for archiving, or from personal contacts/prior knowledge), but the subcommittees should also encourage external proposals, preferably from researchers involved in the original data collection in candidate projects (and thus prepared to contribute to the archiving work) or researchers with innovative ideas about how to re-use the material (and thus prepared to actually carry out the research). Particular weight in the selection might be given to projects that are especially suited for re-use².
- Once a project has been selected, the default strategy should be to encourage *external* reprocessing and in the opinion of the evaluation group this can be accomplished through heavy (technical) support; as judged from the resource consumption per unit (from DDA Sundhed’s perspective), external reprocessing is more cost effective than internal. This seems plausible since the original investigators are the persons who know the details around the investigation and who are best suited to do additional data cleaning and to produce the required

² For example large samples with exposure/prognostic information of current interest, availability of critically important covariates, Civil registration numbers that allows follow-up through record linkages with other registers, and/or an innovative proposal for re-use that is expected to generate unique and important knowledge.

documentation. As a suggestion, support could be in the form of salary for typically 2-3 months for the investigator who is to do most of the practical work, some introduction/training (if needed), and above all close interaction with a designated archiving expert from DDA Sundhed throughout the reprocessing and archiving;

- the basic archiving services, which include the receipt of the material, control of the decipherability of the digital files, and transfer of data, documentation files and a description of the study into a long-lasting archiving format, as well as physical storage of the material, should continue as previously. These services are important because they will insure reasonable care of all relevant research materials, including the rather large proportion that emanates from environments without traditions as regards good data management and which have not reached the critical mass needed for forceful enforcement of good archiving practice. In the opinion of the evaluation group, the interaction between the researchers and the experts at DDA Sundhed could perhaps become even more intense; according to DDA Sundhed's plans, 42.1 hours were allotted to each project (receipt, technical archiving, study descriptions), but the outturn was only 13.5. It seems important that the DDA experts insure that the description of the study, the variable labels and the documentation are adequate and understandable for external reviewers. Thus, the contacts between the researchers and the staff at DDA Sundhed could also be seen as windows of opportunity to spread knowledge and attitudes that will hopefully catch on widely in the Danish health science research community;
- deposited material is to be taken care of without additional delays. The list of deposited materials included a considerable number of materials which had been received but which were not yet archived or under reprocessing. In order to convincingly "market" new materials, it is required that the quality of the materials is assessed, and that the adequacy of the documentation is verified. It would be unfortunate if new materials would routinely pile up awaiting further processing;
- it is important for the research community to understand that archiving at DDA Sundhed is the only way to preserve identification of individuals in materials with Civil registration numbers. If not archived at DDA Sundhed, the Danish Data Protection Agency prescribes de-identification through deletion of the unique personal identification numbers. This will vitiate all possibilities for future follow-up through record linkages;
- DDA Sundhed has developed a nice search function that identifies deposited materials. The results include a description of the study, the status of the material and the availability. It appears that the search is based on keywords. Given that external researchers may be interested in specific exposures/variables that were not the primary focus of the original investigation – variables whose existence may not be evident from the general description of the study – it would be helpful to supplement the searchable material with variable lists, on condition that the terminology can be made reasonably standardised. This might boost the demand for re-usage of the stored material;
- the interest from the international research community might increase if it would be possible to make searches in English;
- it would be helpful if the information about single investigations on the website could be supplemented with links to key publications that describes the material and its potential in more detail;
- information on to what extent different datasets overlap (i.e., cover the same individuals) would be helpful for prospective re-users. If such overlap exists, the same individuals could potentially be followed, through usage of their Civil Registration Numbers, across several investigations;

- the re-use of existing datasets could be facilitated if record-linkages with archived datasets would be done by DDA Sundhed. To ensure the highest level of data protection, the data linkages between archived datasets and other registers could be performed either by DDA Sundhed or the register keeping organisations so that unidentified data are given to the researchers;
- it appears as if the process of releasing stored datasets to external researchers is simple and informal. The person-time allotted in the planning is two hours per material (for preparation of the dataset, registration of the matter, and documentation of the recipient);
- some researchers who have deposited research material at DDA Sundhed have stipulated that re-use is not possible at all. Such restrictions should be sought to be avoided, and furthermore it should be evaluated if these restrictions could be relaxed for instance by virtue of reassuring researchers that all requests for release of the material would be thoroughly evaluated in regard to the relevance, originality and feasibility of the new research, possible ethical issues (conflicts with original informed consent, possibility of backward identification, etc), and possible overlap with other requests for release, ongoing or already concluded/published analyses;
- even though some of the materials can to be used without contacting the original investigators, it would seem prudent, when possible, to refer all applications to the contact persons representing the researchers who originally carried out the data collection in the respective projects for their comment before the final release decision is made.

Hence,

the volume and quality of the DDA Sundhed activities during the years 2005-2007 have been quite satisfactory, both in general and with regard to the specified tasks, although some fine-tuning, as suggested above, might further improve the coverage, inflow and scientific yield;

the organizational solution seems appropriate, but it would probably be beneficial if the medical profession (and/or representatives of various fields of health science research) would be more directly involved in the identification of suitable projects, prioritization of projects for full reprocessing, and decisions about release of materials to external researchers, while the employees at DDA Sundhed would devote more time to teaching/training of researchers and direct interaction with the researchers actively involved in the archiving process. Full reprocessing is a time-consuming task that should be reserved for a limited number of projects where re-usage of data is realistically foreseen. When full reprocessing is considered, it should preferably be carried out by the researchers who collected the data and who are best acquainted with all the details of the investigation. DDA Sundhed should offer support in the form of training, funding, and continuous involvement by a dedicated archiving expert. Internal full reprocessing should only be done in exceptional cases. "Simple" archiving should also involve a fair amount of checking and documentation through close collaboration between the researcher and a dedicated DDA expert, less than in full reprocessing, but probably more than simple archiving today;

the strengths of the DDA Sundhed initiative are so obvious that they hardly need to be mentioned; by promoting good archiving practice (and hence good data management) through teaching, opinion-making, counselling and concrete involvement, the general research environment in Denmark is enhanced in a way that fills researchers in most other countries with envy. In particular, by offering assistance to researchers in environments without traditions as regards good data management and below the critical mass needed for strong

enforcement of good archiving practice, know-how is transplanted while valuable research materials are rescued from annihilation. However, in order to reach the latter researchers, who seem to be underrepresented at present, active “marketing” of DDA Sundhed should be designed so that broad coverage is attained. Repeated exposure in *Ugeskrift for Læger* and *Dagens Medicin* are ways that have been tried, and these attempts should be continued. While historic cohort studies proceeding from administrative data are quite common both nationally and internationally, re-usage of data collected specifically for medical research purposes has been uncommon. Although the factual outturn in terms of ordered and released datasets from DDA Sundhed has, so far, been disappointing, it must be emphasized that it takes time to alter old traditions, and that there is a rather long starting stretch from an idea to actual performance of a study. The existence of DDA Sundhed promises to change the old traditions in the long run. Weaknesses/threats to this scheme are indifference or reluctance from certain important segments of the health science research community to share data. It is noted that clinical research has so far been markedly underrepresented. This is unfortunate since important insights might be gained from follow-up of well-characterized patient materials or large-scale randomized intervention studies (for example the study by Cook et al. [BMJ 2007;334:885] which for the first time linked salt restriction to reduced cardiovascular mortality among normotensive persons). Therefore, extra effort should be targeted at involving these underrepresented researchers. Another threat is potential concern over possible misuse of deposited data, or plain lack of sympathy with the idea of data-sharing, which might prevent potential suppliers from depositing their datasets. Rigorous evaluations of applications for re-use may allay concerns over misuse and prevent media uproar over ethical issues. An obvious weakness would be if the stored and reprocessed material would be severely underutilized and generate little or no new knowledge. It must be realized that most original investigators, no doubt, would like to retain and exploit their hard-earned datasets as long as there are good research ideas, and it is reasonable to assume that the most obvious research opportunities have typically been exhausted when the materials are deposited. Clearly, the utilization so far is unimpressive, but as indicated earlier, it is too early to be seriously worried. It is very unlikely that the entire research community would suddenly run out of creative ideas. However, it is probably important to widely release publications with successful re-exploitation of archived datasets in order to stimulate the thinking of other potential users. Finally, there is a risk that the uncritical use of data by researchers who are unfamiliar with the details of the data collection effort and who fail to recognize the limitations of the investigation might lead to erroneous conclusions. Therefore, it seems prudent to always keep a link open to the original investigators when possible;

as hinted in the foregoing, it is too early to make any evaluation of the cost-effectiveness of the present initiative if scientific progress is the outcome. Suffice to say that at present, the Danish health science community seems to get a lot of “bang for the buck” because the work has generated other values than purely scientific. Given the fairly limited budget, the evaluation group has some concern about the sustainability of the present organization. It appears to work near its capacity ceiling, and with the explicit ambition to increase both input and output, there is a risk that there will be bottle-necks in the production;

as a bottom-line conclusion, the evaluation group clearly endorses a continuation of the DDA Sundhed initiative. However, the scientific utility needs to be carefully monitored, and a renewed evaluation after, say, five years would seem advisable.