
A Danish Research Portal on the Internet

Lately it has been considered to create a stronger web presence for the Danish Data Archives (DDA). In order to do so an analysis of the DDA's current web presence has been made and new strategy for web presence has been suggested. In this article I will try to outline these two steps. The new strategy is based on an ambition of creating a web presence that covers all services carried out by the data archive in a way that supports internal work processes and integrate these with related work processes performed by external parties.

The article will describe the implications for web presence that follows and suggest that web content should be structured around three resource fields named data production, data archiving and data usage. The article is based on a presentation at the IASSIST Conference 2002.

About the DDA and data archiving

The DDA was established in 1973 as a national data bank for quantitative research carried out primarily in the social sciences. As such the DDA collects, preserves and disseminates machine readable research data. In 1993 the DDA became an independent unit in the Danish State Archives. At present the archive has 15 full-time employees.

Traditionally data archives are associated with the social sciences. However, the DDA has never been seen as a resource for the social sciences exclusively. Work areas for the DDA are defined by the potential for exploiting a core competence in preservation of empirical data of a certain structure e.g. a survey conducted as a study within medical science is received just as well as a social science survey is.¹

As colleagues working in data archives know well, it is no easy job to retrieve data from the research community and great insistence needs to be exercised by the archivists. In an ideal world data archiving would be an integrated step within any research project. In reality researchers seldom regard data archiving as part of their research project and this often becomes an unexpected burden when the research project is finished.

*by Anne Sofie Fink**

When data material arrives, data and documentation are converted to an archival format, which secures technical preservation for the future. According to priority the data materials are processed, which implies standardisation and various check of the material. During this process it is often necessary to request information from the researchers in order to make the documentation as complete

as possible.

Dissemination of data material is carried out by providing search catalogues on the Internet that let users search and select materials on their own. Some material will be on-line accessible within NESSTAR² whereas others will need to be ordered from the data archive.

Obligations as data archive

As the national data archives for social, medical science and history, the DDA has an obligation to act as an active partner in the Danish research environment centred on empirical research. The archive's contact and services to external parties is in this respect largely dependent on the web. Among other things this means that www-based facilities for interacting with the research community continuously must be adapted and developed.

Although Danish researchers and students will be perfectly able to use a web site in English, one aspect of this obligation is in my view to supply DDA's Internet service in the national language, Danish. To national producers and users of data, a web site in Danish will call attention to the DDA as an active player in the national research environment.

Our current web presence

It is frequently pointed out that a lot of organisational web sites are internally focused in the way that the sites are much more about describing the organisation, than about offering information relevant to outside stakeholders. The DDA web site is currently no exception. The web pages are used as if they constituted an information folder. They are static, there are no links to other sites and visitors are met with long 'dead' texts to point out a few examples.

Besides the web site the DDA supplies: a search catalogue

on the archive's holdings, a search catalogue which is part of the Integrated Data Catalogue (IDC) and a NESSTAR-based catalogue. This article will not make an evaluation of these search facilities, but – for obvious reasons – they are exclusively devoted to search and location of data sets.

Figure 1 below shows the relation between the data archive and its stakeholders and web presence.

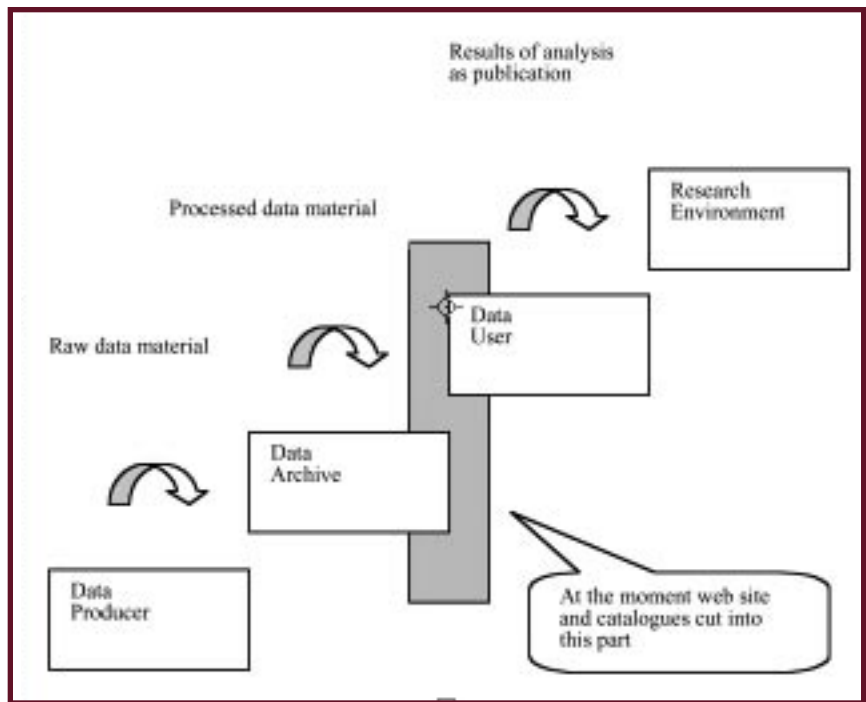


Figure 1

An interpretation of this figure shows that the web site and search catalogues mirror to a simplified production line where dissemination of data is the only activity given visibility on the web. Some important insights can be gained from making an elaborate interpretation of the figure 1.

□ Actors

There are a variety of actors present. But these actors may in real life perform different roles – the producer may become a user of data, the producer may become disseminator of data, the user may become a producer, etc.

□ Flow

From the model linearity is presupposed, but more arrows could be made since both data sets and communication may be seen

to move back and forth among actors and sometimes jump the actor first in line.

□ Universalism

The model which is used seems universalistic in scope, but any research environment will be unique in many ways e.g. due to the unique national context and it should be possible to mirror this on the web.

□ Limited transparency

Web site and search catalogues cut into the 'travel' performed by a data material by focusing only on dissemination. However, all parts of the travel are relevant not only to the data archive but also to external parties. The web service should support all activities performed by the archive, not to create a duplicate existence to the DDA on the web, but to create a bunch of complementary activities implying great synergy effects to be gained.

Another perspective on web presence

The interpretation of figure 1 suggests viewing the data archive not just as collector and disseminator of data but also as an intermediary agent between data producers and data users (and disseminators of knowledge products). As intermediary the data archive would enable flows of communication and data sets among actors who will be taking on interchangeable roles of data producers,

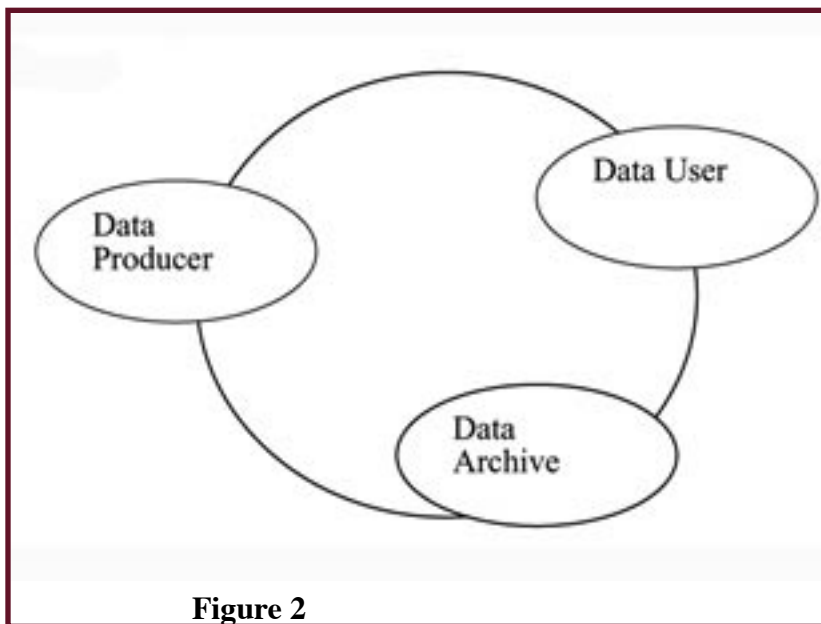


Figure 2

data archive and data users. Therefore a network model seems more fertile in showing relations between data users, data archive and data producers, see figure 2 below.

As roles are interchangeable I suggest giving up viewing them as separate entities to be targeted and instead complementing the network on the web by three resource fields – a field for data production, a field for data archival issues and a field for data usage. In this network the data archive would act as the organiser of information content leading users into a virtual space structured around the three fields. This would create a broad coverage of subjects related to use, storage and production of data sets not as closed rooms but linked in a way that supports the visitor’s work where – by way of examples – issues concerning secondary data will be intervened with data production and preservation. This would be to supply an integrated information gateway that supports non-linear work processes and communication flows among actors.

By constructing the web site around three resource fields, the web service will incorporate great flexibility. The strategy would be to supply a bundle of content related to each field, which under goes continuous construction and adaptation according to visitors’ needs and demands. Thereby the service will be tailored to the national research environment the data archive is part as an ever-changing reflection and support ‘organism’.

With this broader perspective on services provided on the web, the data archive should not only be seen as a service towards data users but just as well to data producers and data archivists since the service will be addressed both externally and internally. External services to data producers will have the aim of opening up the black box the archive has been so far. For instance when a data material is handed over to the archive, the data producer is no longer part of or aware of the work processes taking place within the archive. What will be done in this respect is to turn the archive inside out. Closely linked to this ‘turning inside out’ is that the services should

include services supplied to the data archivists themselves e.g. search facilities for administrative purposes, standards for documentation and networks for corporation to name some.

The data archive as organiser will not define what is relevant for users of the service/visitors of the portal instead this will be put up to them to suggest and add content. What the archive will provide is management of structuring the content and where relevant supply of information content. In this way the same navigation rules will be place on internal and external users of the fields.

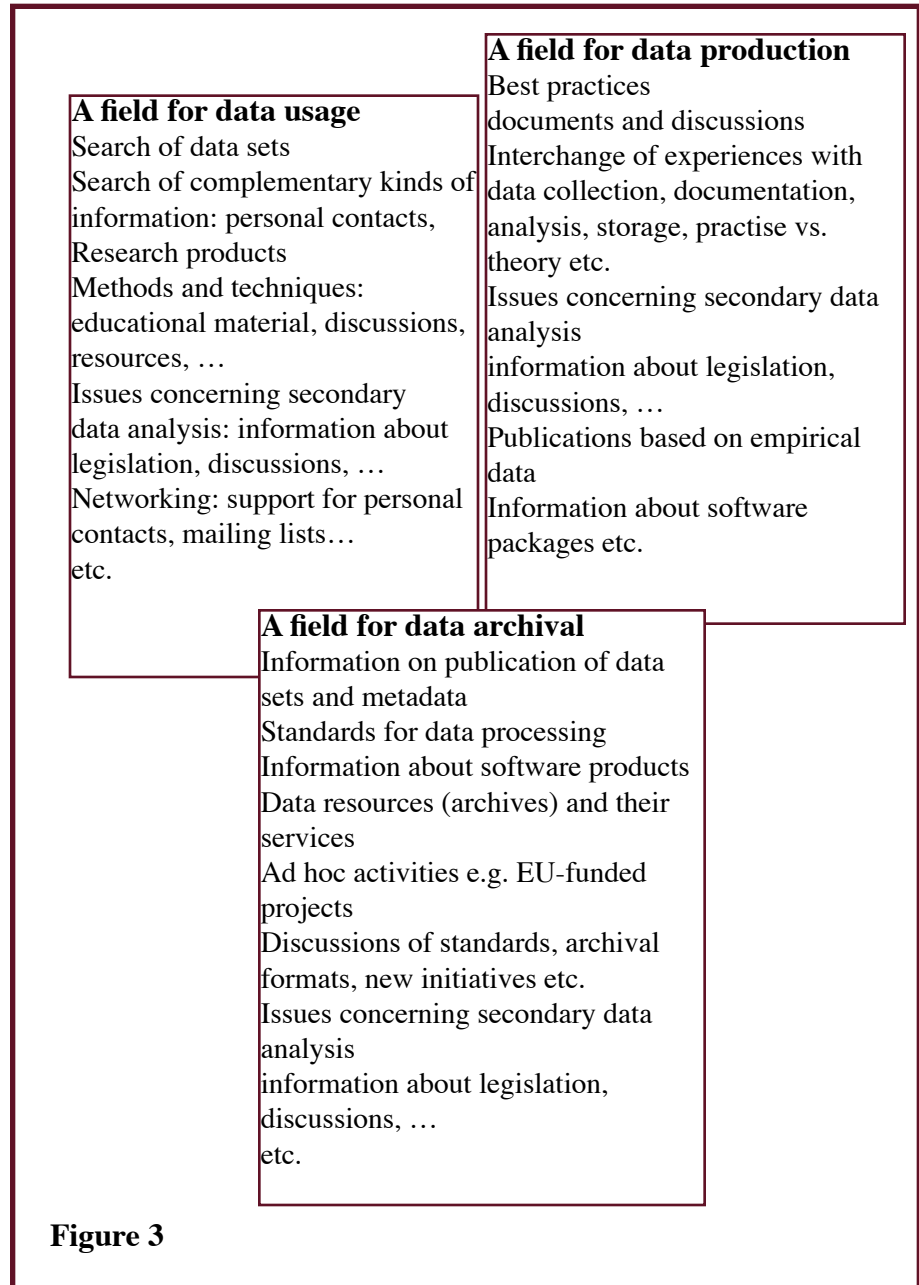


Figure 3

In order to conceptualise this kind of expanded web service supplied and managed by the data archive the word web portal seems more appropriate. Through this research portal data producers will be able to check issues to be aware of when carrying out a survey, data archivists will be reminded about procedures for data processing, users will find data sets across Europe to throw light on a problem, etc.

Content of the resource fields

In figure 3 a selection of ideas for content for the three different resource fields is listed. As visualised by the figure the three fields share some aspects whereas others relate predominantly to one of them. To clarify the argument for the resource field view I will comment on a few items related to each of the fields.

Search for data sets and more

To a resource field for data usage obviously there must be links to facilities for searching for data. An excellent example for this is the NESSTAR system, which facilitates search, data and metadata browsing, on-line access to data analysis, graphic presentations and downloads. Ideally there should be no limits to the search engines offered. The challenge is to provide users with an efficient gateway to this a vast amount of engines. One feature would be to make users' evaluation of different engines available to all of the users.

Users searching for data will seldom have an interest limited to the data and metadata provided. They will often have an interest in complementary kinds of information such as articles based on certain data sets, experience in analysing the specific data set, etc. (as is acknowledged by the MADIERA project³), too. But moving in the opposite direction 'back' to the data archive and producers is also relevant (see figure 1). For instance the user might need to know about the principles for data processing used on a material or might want to know if it is possible to retrieve additional documentation by approaching the data producer personally. In this way the portal will acknowledge that access to data is only a part of what the data users need and the portal will support this.

In this respect a resource field for data usage will in some respects be intervened with the two other fields, although the data user will experience this as integrated discoveries within a search process.

Data Processing

On the surface a field for data processing seems to be relevant only to data archivists – a view which some times is carried forward by data archivists – but it is actually an issue that concerns all users. To give an example, users will often need to become familiar with standards for data documentation, since they will influence the quality and/or sophistication of the data analysis the data material

will support. To carry the example further, in order to accommodate this, standards for documentation have to be acceptable to the data producers and the data archive alike. The implication of this is that the archive needs to gain support and acceptance for its documentation accepted by the data producers in order to make them willing to supply the necessary information (and ideally without the data archivist having to make explicit demands for it). The field for data archiving could provide standards for data documentation applicable to data production and data archiving simultaneously.

Although users seem to have a great hunger for documentation, inflexible demands for elaborate documentation put on data producers might lead them to refuse to offer their data to the archive. Transparency and communication mediated by the research portal are expected to lead to greater understanding among actors.

Data production

As it is stressed several times now by way of example the expectation is that a field for data production is not a source only relevant to data producers, primarily researchers, but also to other visitors of the portal. For instance legal issues concerning conduction of a survey is something the researcher must be informed about. But also to the data archivists certain legal issues are important to be aware of when accepting to store, process and disseminate survey data. To users of the data it will also be important to be aware of the legal restriction a survey is conducted under in order to integrate this background information into the analysis process and conclusions.

Personal contacts

An often-celebrated feature of the Internet is the possibility of connecting people across space and time; the portal must obviously support this. Discussion forums and mailing lists are aspects to take advantage of in this respect. As an archive concerned with adapting to visitors' needs and demands, virtual forums may be used to pose and discuss questions e.g. discussions of standards, archival formats, initiatives, etc. The support for personal contacts and discussions obviously reaches across the different fields and will act as a tool for defining and providing information content for the portal.

By elaborating these few examples the argument for a portal structured around resource fields has been made. Hopefully, the goal of supporting work processes and communication that does not fit in with a traditional strict division of usage, storage and production has been detected. To conclude; whether one is conducting surveys, collecting data, storing data, performing secondary analysis there are interrelated issues to be concerned with. A portal structured around the three resource fields will be able to represent an issue as a whole but offer multiple entries to the visitors based on his/her immediate preferences.

Managing a portal

In order to supply a portal in line with needs and expectations of the three target communities – users, archivists and producers – it is essential not only to collect feedback, but also to collect it in a way so that is readily available for evaluation and implementation. The feedback then needs both to be structured and to be discussed openly. This means that feedback will be collected and put up on the web site to prompt discussion among users from all communities.

For a small data archive the workload involved in constructing and running a research portal might seem overwhelming. In order to meet with the challenge put on the archive, two fundamental principles should guide the work. First of all, external parties must be involved in the work and teamwork between external and internal parties will be essential. External parties' involvement will admit access to knowledge, insights and artefacts, which is not held by the data archive. The second principle is implementation of a modular structure of content. It should be possible to add, change, remove, and adapt content step by step and thereby incrementally construct – and de-construct – the portal content. In that way the portal should never be seen as complete or finished. What is essential however, is that strict structural guide lines are implemented in order to make sure that content are accessible in a well-organised and systematic way. This is also to make sure that the portal stay 'manageable' to the data archive as the key actor, otherwise the objective of supplying content with a high degree of usability will slip out of hand.

“Accelerating Access”

2002's theme for the IASSIST Conference was “accelerating access” and this article was in line with this theme. However, my goal has in some ways been broader than that. With the resource field approach the data archive is expected to open up to the outside by supplying a well-structured information gateway with access to a broad range of topics concerning production, storage and use of data.

On a micro level the portal will offer a multipurpose tool for supporting people working with empirical data. On a macro level the portal is supposed to raise awareness and use of secondary data as well as data archiving in the social sciences and other scientific disciplines – potentially also to other areas such as consultancy work and public accounts.

Last but not least it is our hope that the portal will give greater visibility to the DDA as a competent and experienced actor on the national research scene and thereby support the activities carried out by the archive.

Footnotes

1 From 1996-2002 the DDA received a separate funding

for registration and archiving of data from the medical sciences. This initiative was named ERAS - <http://www.sa.dk/dda/eras/english/eras/default.htm>.

2 NESSTAR – Networked Social Science Tools And Resources – www.nesstar.org – is a web-based system for searching, browsing, analysing, presenting and downloading statistical research data. The DDA is one among several data archives distributing data through NESSTAR.

3 The MADIERA – Multilingual Access to Data Infrastructures of the European Research Area – project picks up where other NESSTAR-centred projects have stopped. The project objective is to create a multilingual search engine that gives users access to a broad range of data materials and related materials.

* Paper presented at the IASSIST Conference, May 2002, in Storrs, CT, USA. Anne Sofie Fink, researcher, Danish Data Archives, asf@dda.dk.