Plugging the digital gaps: the use of metadata in the Swedish Social Insurance Administration

Roger Löfgren and Lars-Erik Hansen

Dr Lars-Erik Hansen is Chief Archivist of the TAMArkiv (The Swedish Archives of the White-Collar and Professional National Unions) and **Roger Löfgren** is Senior Archivist of the Swedish Board for Study Support (CSN) in Sweden. They were responsible for the Archiving Project at the Social Insurance Agency in Sweden. Lars-Erik Hansen has also worked at the National Archives of Sweden in the Division for Inspections and Consultation. Lars-Erik has a doctorate in history from the University of Stockholm. Roger holds a master's degree with a major in ethnology.

The theoretical discussion on digital archives has been extensive for more than a decade, but there are few examples of practical and real implementations. During the last five years, the Swedish Insurance Agency (SIA)[1] has implemented a digital archiving system to manage the tens of thousands of documents produced daily in the case management systems used at local social insurance offices. The system was tested in a pilot version and fully launched in 2006. The digital archive at SIA works, however continuous evaluation, feedback and development will improve and expand the product. The digital archive project has found the OAIS, ISAD (G) and EAD standards to be usable and good, but also found that they cannot be implemented too rigidly. Instead, a more flexible approach is needed, using the standards as a resource whereby relevant metadata is added along the whole document life cycle.

The aim of this article is to describe one of the biggest digital archiving projects in the Nordic countries undertaken in the period 2001–2006. It was run in one of Sweden's biggest public agencies: the Social Insurance Agency in, collaboration with the National Archives of Sweden. The major part of this report covers the period 2001–2006, although a few changes have been made during 2007–2008.

Background and pre-conditions

Researchers generally believe that the Nordic welfare states offer their citizens the world's most fully developed social security systems – from cradle to grave. The Swedish Social Insurance Agency is the authority which is charged with administering the welfare state. The agency makes decisions on the right to allowances and makes payments, from children's allowances to pensions, for all Swedish citizens.

Toward the end of the 1990s, the Swedish government formulated a vision for e-government with the intention that all information and service that could be provided electronically would be so provided, as long as it was done cost-efficiently. The vision was motivated by the need to make as efficient a utilisation of society's common resources as possible through the medium of new technology while providing increased access to the welfare state's services for the citizen. The Swedish government's intention was that the citizen would be able to choose how to approach public services – a face-to-face meeting, writing a letter on paper, by telephone or directly via the Internet. This increased accessibility via various channels would also facilitate collaboration between authorities and companies that are service-providers.

The Swedish Social Insurance Agency's case management system (CMS)

Given this framework vision for e-government, the Swedish Social Insurance Agency began an extensive development project, the aim of which was to produce an electronic case management system (CMS). In 2002, an initial version of the CMS was put into service, encompassing 2.5 million cases per year.

The easiest way to describe the CMS is that it is an electronic document management system (EDMS) which handles a number of administrative

functions. The system supports centralised scanning of paper documents relating to an application (or case). Electronic applications can also be submitted by the applicant via the Internet. An e-reception desk supports the distribution service to send the application to the correct case manager for supplemental metadata augmentation, inspection and decisions on refusal or acceptance, which may result in a payment. Payments generate a request to a separate payment system. The documents are stored in several different formats according to an internal Social Insurance Agency (SIA) standard and are held together as a virtual case file with the aid of references (metadata). In addition to document management, the case manager fetches and sends data to and from databases other than the CMS and from other agencies, organisations and employers. In later versions of the CMS, automatic case management has been implemented with the support of a rule engine and use of simplified communication between the case management in the system and adjacent databases.

Conditions for destruction and preservation from CMS

In order to further facilitate the understanding of digital preservation within the SIA, the reader needs a brief description of the Swedish rules for preservation and destruction of records.

Records become public according to law when they arrive, are processed or otherwise completed, and are stored, by a public authority. Furthermore, according to Swedish law, the set of public records kept by an authority make up its archives (fonds),[2] based on the principle of *respect des fonds*. An authority's archives shall, according to law, be cared for, preserved and kept separate from other archives. The archives shall also have an archival description. Public records are, by old Swedish legal tradition, in the public domain and it is important that, apart from internal operational needs, they are retrievable, comprehensible, accessible and preserved for research from the perspective of public access and democracy. Before a record is made accessible to the citizen or the researcher, a secrecy assessment is performed according to an extensive body of regulations.[3]

Drafts (working versions) for public records may be destroyed without regard to any special regulation, while the final version must be preserved. Public records in an authority's archives, however, may

only be destroyed in accordance with the regulations of the Swedish National Archives, unless legislation other than the basic archival legislation takes precedence for preservation and destruction (for example, The Secrecy Act). When electronic management in the CMS for Social Insurance Agency benefits was introduced, just such other legislation[4] took effect.

At the time the new SIA system was developed, a new law regarding privacy issues for individuals was introduced.[5] Previously there were destruction limitations on the paper records that made up a case. With the new law in place records were to be destroyed in their entirety but the law left room for the National Archives in consultation with the SIA to produce regulation aimed at preserving these electronic records.

The new regulation was formulated on the principle that electronic records pertaining to a case would be preserved in the same manner as if the information had been produced on paper. The focus was then on the preservation of electronic documents that make up the cases and on describing the case handling according to an archival description based on ISAD (G)[6] and ISAAR (CPF).[7] Information in adjacent databases that is used in the CMS would be preserved using a separate procedure. The Swedish personal identity number, assigned to every Swedish citizen, was to become an important reference between the preservation of documents in cases in the CMS and information in the adjacent databases. Information in the databases was preserved in the conventional manner, through exports to flat files directly to the National Archives, without passing through the SIA's long-term archive. Plans exist to record information from these databases in XML documents and archive these in the SIA's long-term archive, but this has not yet been effected.

The strategy and the project for digital long-term archiving

The archivists of the SIA preserve and care for the archives, make them accessible and have a central function to provide support in archival matters within the authority. Since the end of the 1990s, an ever-increasing amount of information is handled digitally. Digital information management requirements have therefore strengthened

and developed. A strategy for the digital preservation of information was produced in 2001, early in the project. It contained 12 items:

1. The ability to read from and search in a long-term digital archive must be independent of the systems and applications that originally created the records.

Records are created by many different systems and applications, for example word processors such as Microsoft Word. In order for archived records to be able to be read and found, these records needed to be transformed and stored in a format that is archive friendly. That is, we expect software to be available to understand the file in the very distant future. This means that the various formats that were specific to given applications could not be used for archiving, and the data contained within the format needs to be converted to a more archive friendly format.

2. All archives within the SIA must adopt the same storage structure and standards in the long term.

Within the SIA there were different structures and solutions for managing archives. The project members decided it would be more efficient if the same solution were used overall, since the control of information, regardless of the medium (digital or paper), then became more uniform and easier to understand.

3. The files are stored in the original system (the near archive) until the case is closed. Then the files are moved to a long-term archive.

4. A selection of archival records are to be transferred for research purposes to the National Archives after an agreed interval.

It was important to build up routines and procedures that ensured regular transfer of records that were to be preserved by the National Archives.

5. Metadata that is to be stored should be created as early as possible.

Metadata is created and used during the entire lifetime of a record. Some metadata was relevant only to long-term archiving.

6. Records in a file produced in-house are to be stored in XML.

In cases where the SIA itself determined the format of a record, for example for journal notes, XML was to be used. The reason for this was that XML is application independent and is presentation neutral.

7. Final versions of records created in Microsoft Office or other third-party products and that are included in the file are to be stored in XML. If this is not possible, the record should be stored in PDF/A.

Third party products such as Microsoft Word underwent constant change and new versions were regularly introduced. It was important to choose more durable formats such as XML and PDF/A. As has been mentioned above, XML was preferable.

8. Information that is subject to long-term preservation should be created in the format in which it will be stored, as early as possible in the process.

Whenever possible, durable formats were to be used.[8] If documents are created and stored in a long-term format as early as possible, costly conversions, version problems and possible data loss can be avoided.

9. Where relevant, the OAIS[9] model is to be applied.

The OAIS model is considered to be an abstract, complete model for archiving systems, which described at a high level the various functions, processes and information structures of archiving systems. Furthermore, the OAIS model introduces a powerful general conceptual framework. By using this model, it was unnecessary for the SIA to 'invent' a model of its own. The model has also provided a tool for the SIA archivists in their efforts in setting requirements, for example, in the question of metadata.

10. ISAD (G) and the EAD[10] standard are to be used for archival description and to facilitate searchability in the digital archive.

11. Digital signatures at the record level are to be replaced by a 'stamp of authenticity' directly on receipt.

In order to avoid all further dealing with digital signatures, these were to be removed as early as possible. Digital signatures cause problems in the long term, since certificates were often time-limited and it was also usual for records to be encrypted and signed. In the worst cases, it would be impossible to decrypt a record. It was, however, important for the SIA to guarantee the authenticity of a record, and here alternatives to the original signature were produced.

12. Archival descriptions, flows and deliveries should be automated to the greatest extent possible.

In order to reduce the cost of constructing electronic archives and transfers to the National Archives, processes were to be automated to the greatest possible extent.

In addition to the 12-point strategy, a decision was made between various long-term storage strategies. The possibilities considered were:

1. printout to relatively stable media, for example, paper or microfilm;
2. total preservation of both hardware and software;
3. emulation of hardware or software in original format; and
4. migration or continuous conversion to current, stable formats.

The archiving project evaluated the various strategies. The choice fell upon migration, or a series of conversions to be carried out over a period of time. This strategy also required continuous administration, but it was judged not to be as extensive as the administration of a total preservation or one which involved emulation. Even if migration would certainly require continuous development, it was judged to be considerably less than would be the case if an emulation strategy had been used. The loss of information was expected to be less than with a printout strategy, since retrieval capacity did not diminish as much as with printouts.

After this strategic work, extensive labour with demand management began, which deepened knowledge of digital preservation and shed light on the problems that can arise in the absence of a long-term archiving strategy. The strategy formed the basis for continued efforts.

Strategy and requirements put into practice

When the strategy had been produced, a sub-project dealing with digital archiving was started within the larger CMS project. A requirements specification was produced, with functional requirements that were structured on the basis of the OAIS model. The relationships between existing and planned systems were illustrated as shown in figure 1 on the next page.

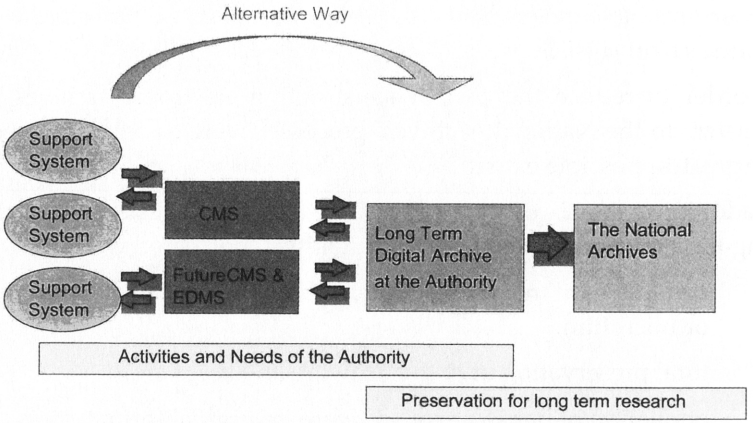

Figure 1. Information flows between different systems

According to the model, applications and supplementary information received from clients were input via support systems for processing in the CMS, such as scanning, e-reception and the distribution system. Adjacent databases were also regarded as support systems. Information was fetched from these databases and sent for case processing in the CMS. The model included support for future CMSs and document-based information systems, for example future systems to support administrative operations within the SIA. The SIA also needed to provide for long-term archiving of cases in a more inactive system (long-term archive) for reasons, among others, of efficiency and security. The ability to retrieve cases, or information on cases, from the long-term archive to the CMS is essential. The same need was judged to exist for future CMSs and document management systems. Also, when the original systems were ready to be de-commissioned, efficient and more user-friendly search interfaces must be produced for archived information and maintained for as long as it was needed in operations. Such search interfaces could be connected both directly to the long-term archive, and to new CMSs and document management systems replacing the old ones. When operations no longer have a need for a case, it should be possible to extract it from the long-term archive and destroy it. Cases that were to be preserved for research at the National Archives were to pass over to their archive system (due to the new, stricter integrity

protection legislation previously discussed under 'Conditions for destruction and preservation from CMS'). This transfer of sample cases had to happen no later than the point at which other cases were scheduled for destruction. The archive system of the National Archives was furthermore judged to be in need of a search interface where case information could be searched and visualised in a manner that would provide a satisfactory substitute for the original CMS, which was deemed impossible to preserve in its entirety in accordance with the migration-and-conversion preservation strategy that had been chosen.

All transfer of information was to be carried out in an organised manner by means of agreements between the sender and the recipient of the information. Against this background, the system owner of the CMS needed to place demands on the support systems in order to get the right information at the right time, just as the system owner of the long-term archive needed to place demands on packaging and identification of information that was to be archived. The National Archives in turn placed demands on the material delivered from the long-term archive at the SIA so that research could be facilitated. The demands of the National Archives were then reflected in the demands on information from the CMS to the authority's long-term archive. Within the OAIS concept, this line of reasoning is represented using the following concepts: *submission, archival* and *dissemination information packages* and the functions of *ingest* and *access*.

The 'alternative way' indicated in figure 1 refers to the separate transfer procedure from the databases used for processing in the CMS, but whose data was not recorded in it. The case worker sometimes had to search for additional information stored in the databases when the source of important information was not well documented in the case files. Selection for preservation from these databases was made in the conventional manner through withdrawals in flat files, without passing through the SIA's long-term archive.

The OAIS model put into practice – interaction between producer and consumer

The strategy required that the OAIS model be applied wherever relevant. Based on the requirements specification for the long-term archive, information was to be submitted, archived and distributed in terms of *information packages* in accordance with OAIS-style *agreements*. Each case would be considered as a submission information package (SIP) on arrival at the long-term archive and as an archival information package (AIP) on archiving in *archival storage*. The project group chose that the *SIP* be supplemented in the *ingest function* with metadata on preservation and destruction. Communication between the *producer* and the *primary consumer*, the CMS, was adapted from the OAIS figure 2.4.[11]

Destruction is not dealt with in the OAIS model, nor does it contain a standard for metadata, but OAIS does contain recommendations on certain classes of preservation metadata or preservation information that should be included in an information package.

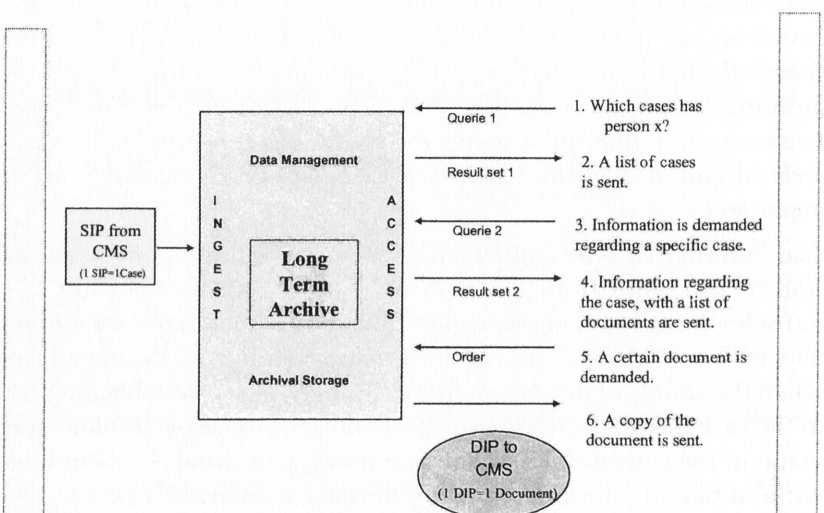

Figure 2. Communication between the CMS and the authority's long-term archive

Metadata in an OAIS archive – the imperative of digital preservation

A word that is often used in discussions on digital preservation is metadata. Metadata is a prerequisite for all structuring, understanding and retrieval of information. There are a number of standards for metadata. Those who worked in the SIA's digital archiving project realised that a framework needed to be available, a structure that was flexible. OAIS provided such an overarching framework.

In OAIS, metadata is dealt with within the *data management function*. Here, the preservation description (metadata needed to understand archived information in the long term) is coordinated for the archive information packages that are stored in the archive. OAIS provides a rough framework for the kind of preservation information that is needed.

Since OAIS does not in itself constitute a standard for preservation information or metadata, it was initially decided that a number of selected standards would be prioritised in order to produce relevant preservation information for the archive packages: Dublin Core,[12] ISAD (G), ISAAR (CPF) and, for transfer to the National Archives, EAD and EAC.[13] At the same time, it was found that metamarking of the information is time-consuming, so requirements for far-reaching automation formed an important starting point for this part of the project. The efficient administration of metadata was also considered a critical issue.

The SIA's document management standard, eDoc

The SIA's internal document management standard for the distribution and retrieval of electronic documents within the CMS played an important role, particularly in scanning paper documents and in electronic forms for e-services. Initially, the SIA looked at Dublin Core as support for this work, however as work proceeded this standard was abandoned because it was not sufficient for the SIA's needs. The internal document management standard that was produced came to be abbreviated as eDoc, from electronic uniform document.

The eDoc standard was and is divided into three levels:

1. metadata about the document itself (chiefly information on the scanning process, e-reception and distribution to the right case manager and case);
2. metadata about the person or organisation with which the document deals; and
3. data on the SIA's consideration of payment within a given benefit.

Certain metadata is obligatory and other metadata is optional. An example of obligatory metadata is data needed to link the electronic document to a person or an organisation if it is external; and to identify the right type of benefit case (sickness benefit, parents' allowance, rehabilitation, and so on).

When a specific form has to be produced for handling a case through the CMS, a document type definition (DTD) based on the eDoc standard was ordered for this form type, which was then followed to create the individual electronic documents or to create an XML string for scanned documents in TIFF format. Journals for cases (summary histories of what had happened in a case) were stored in a single XML file for each case. Exclusively electronic records, for example, applications submitted via forms on the SIA's external website, were also stored in an XML file. Scanned paper documents were stored in TIFF format but their description was compliant with the eDoc standard in XML, which required inclusion of metadata that ensured they were forwarded and attached to the correct case. The electronic documents were stored in one or two data files: in an XML file or in a TIFF, or PDF file with a descriptive XML file.

SIPs from the CMS to the long-term archive

For transfer to the long-term archive of electronic cases with documents compliant with the eDoc standard, a number of demands were placed on file formats and metadata. The file formats that were accepted were TIFF, PDF and XML. PDF/A was not implemented in the 2001–2006 period, but the project recommended PDF/A as a long-term solution. The members of the project demanded that all format handling have active administration and a conversion and migration plan.

SIPs from the CMS were given a content component and a metadata component in accordance with the OAIS model. Each individual case was considered an integral SIP. Each SIP was given an identity (in the package description) and a list of which files it contained (in the package information). In order to be able to read stored file data, file specifications were needed (representation information, not represented in the illustration). For the XML data files, a DTD that complied with the eDoc standard was also required.

The content was contained as shown in figure 3 in the electronic documents stored in PDF, TIFF or XML. Each PDF and TIFF was described by an XML file. In these cases, the XML file contained metadata about the information, not the information itself. For XML files that described other files, only eDoc's data levels 1 and 2 are used, that is, distribution metadata, and search and identification data, respectively. eJournals and eForms entirely in XML contained both content and metadata. Data levels 1 and 2 contained metadata and data level 3 the actual content.

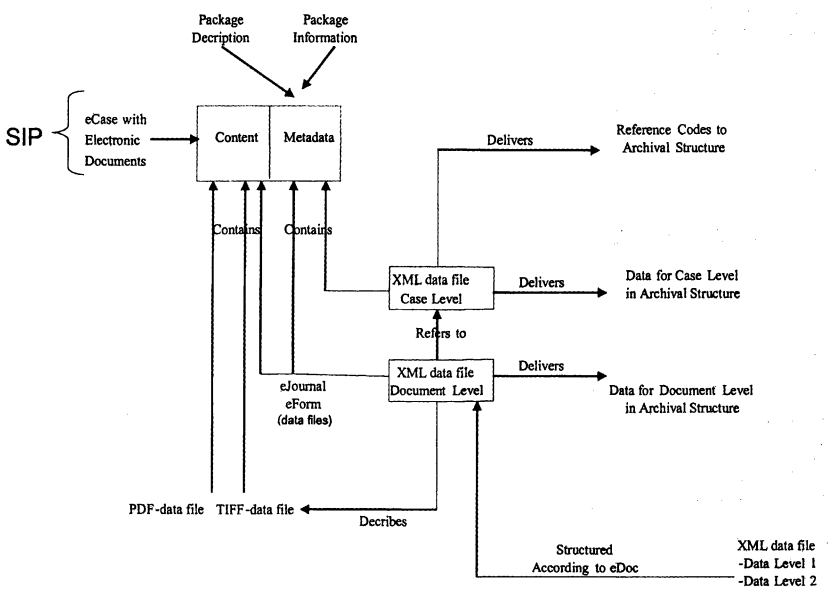

Figure 3. SIP from the CMS to the long-term archive

For the case level, a separate XML file was included in the SIP. This file referred to the respective XML file for the case at document level and also contained reference codes that linked the specific case to the right levels in the archival description structure (fonds, series, sub-series). The information in this case-descriptive XML file was transferred to the archival description's case level. Information from the document-descriptive XML file was placed in the archival description's document level. The XML files, as they were when they arrived from the CMS, were stored in the long-term archive in their entirety. The PDF and TIFF files were also stored as they were when they arrived at the archive.

Application of ISAD (G) and ISAAR (CPF)

For as long as cases were stored in the CMS (in its near archive), case managers could search and read the case and appurtenant electronic documents directly through the CMS. Searches for cases that had been transferred to the CMS long-term archive were performed as shown in figure 2. From the case managers' point-of-view, searching for cases and documents in the long-term archive was similar to searching for those still stored in the CMS. The metadata that case managers most often used for searching were the client's security number, the time period, the type of case and the document type.

The SIA used ISAD (G) and ISAAR (CPF) to produce a description and search structure for metadata about archived information that met the needs of the long-term archive, as shown in figure 4 below.

The ISAD (G) standard contains both levels of descriptions and elements of descriptions. The SIA chose five levels to describe its activities' archive. The first level was the authority level and the archive that is to be kept separate in accordance with the principle of *respect des fonds*, which also corresponds to Swedish legislation. The second level was a first grouping of the areas of activity that were identified within the SIA. These include the areas of pension, sickness and benefits for families and those with disabilities. The third level described case management processes within the areas of activity. The fourth level described the individual cases in which the various case management processes resulted. In the model, the case level was equated with 'file' according to ISAD (G) terminology. The fifth level described the

the individual documents that are linked to the individual cases. In figure 4, the document level is equated with 'item' according to ISAD (G) terminology. Then the elements that were necessary for a sufficiently good description of each of the levels were identified. Reference code, title, dates, creator, and scope and content were considered necessary at all levels.

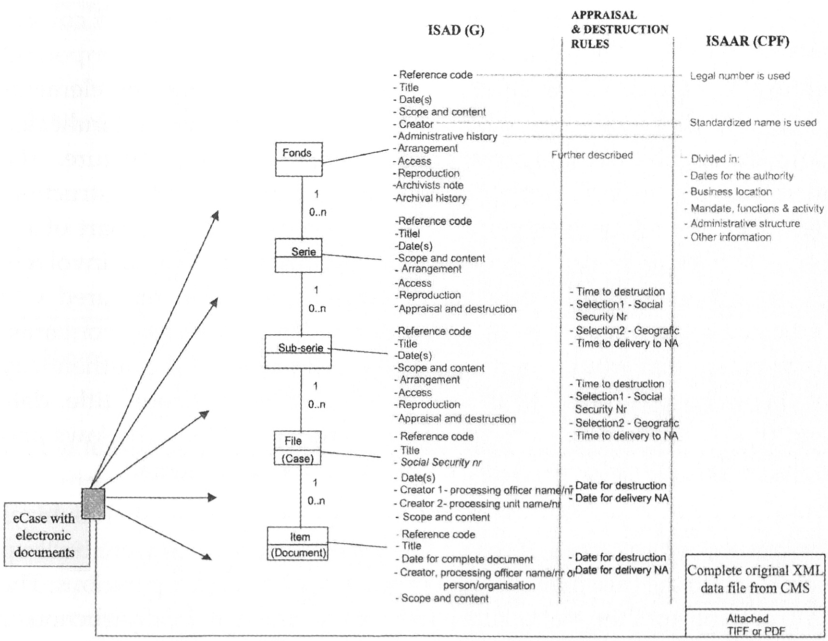

Figure 4. The Social Insurance Agency's implementation of ISAD (G)

The ambition of the long-term storage strategy was that the metadata should be documented as early as possible. The metadata capture for SIPs was limited to reference code, title, dates, creator, and scope and content. For the fonds, series and sub-series levels, capture of reference codes was required as a minimum. For the completion of information for the other elements of description at the fond, series and sub-series levels, an entry and display program linked to the *data management function* was developed for manual entry by the archivist (see figure 5). For longer texts in the archival description, this was seen as the only

practical route. However, for briefer information such as title and dates for the fonds, series and sub-series levels, a strong wish remained that these would have been captured and included within the SIP.

At the fonds level, the legal number for the SIA as an authority was given and, in the title and creator element fields, the standardised name form (the name of the archive became the same name as that of the authority that had created the archive). The dates when the authority was founded and when it may be terminated were given as dates. Scope and content provided information on the content of the archive, and more important groups of records. Administrative history was divided into sub-elements in accordance with ISAAR (CPF). The system of arrangement indicated how the archive was organised, including the series structure. The metadata for 'access', 'reproduction' and 'appraisal and destruction' described the regulations pertaining to these areas. On the part of the SIA, some references to laws, ordinances and instructions were involved. By whom and how the archival description had been prepared was indicated in the 'archivist's notes' field, for the 'archival history' contained information that would contribute to the assessment of the authenticity of the archival records. At the series levels, reference code, title, date and information on rules for destruction was provided. The laws and regulations applicable to cases in the series were documented.

The rules for destruction and delivery of case files to the National Archives for research were placed at the sub-series level. Sub-series were built up according to the case management processes followed in operations. The instructions from the National Archives were aimed at the destruction or preservation of the individual cases containing documents about persons (marked with social security numbers) that were created in the various case management processes. The case process illustrated in figure 5 is for persons for whom compensation for sickness and inability to work has been considered. The rules for destruction and preservation for the cases in this process were: destruction 10 years after closing of the case, with the exception of a sample delivered to the National Archives for persons born the 5th, 10th and 15th of each month, and for persons whose records were processed by certain regional units. The project required that a new or changed rule must be verified for safety's sake by another archivist than the one who determined or changed the rule.

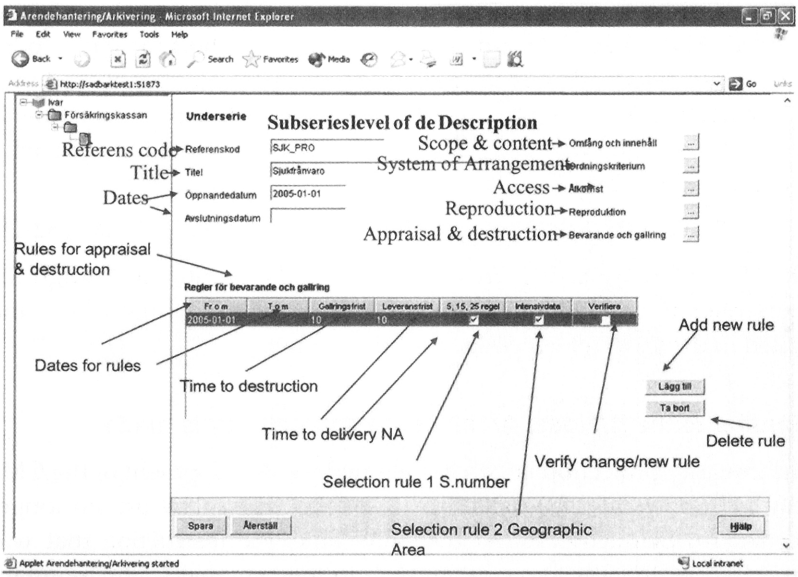

Figure 5. View from the register management program in the long-term archive, sub-series level of description

At file level (see again figure 4), a unique reference code accompanied each individual case and the title constituted the name of the case. Dates were divided into a start date and an end date for the case. Creator was further divided into two elements: social security number, which was a completely determinant reference for efficiently finding the case, and other information on a given person.

At the item level, there were five elements of description within the archival description framework. In 'dates', the time the document was completed and thereby became public according to Swedish law was provided. In archival storage, according to OAIS, the original files sent from the CMS, that is, the TIFF, PDF and XML files were stored in their entirety in compliance with eDoc. Thus it would be possible to search on more metadata than that contained in the archival description structure set out in figure 5.

Although the figures do not illustrate this fact, each reference code was divided into a short and a long reference code. The short reference codes were those described above, for example, the legal number at the fonds level and a unique case ID at the file level. In order to link the documents at item level all the way up to the fonds level, there was a long reference code. This could be:

202100-5521/SJFOR_ÄHS/SJK_PRO/63221313572/494303282929

The long reference code started with the fonds, continued through the short reference codes for series, sub-series and individual cases, and ended in the short reference code for the document.

Transfer to the National Archives – a test delivery is made

The integrity protection law that governed the development of the CMS requires immediate destruction of case files when they are no longer needed for SIA activities. At the same time, the instruction that was produced by the SIA in consultation with the National Archives entails that a sample of the cases are preserved for research. The first delivery of cases for research must take place no later than 2013, since the first cases that have been processed via the CMS will then no longer be relevant to the SIA's own needs. In order to test the new concept of their digital long-term archive, the agency contacted the National Archives as early as 2004 to perform a trial delivery. A formal request was made to deliver cases from the CMS via the digital long-term archive to the National Archives. As a result of this request, the SIA was used as a pilot case for the National Archives' project for long-term digital preservation (the LDB Project).[14] This project started in 2004 and was a collaboration project between the National Archives, the Luleå University of Technology, the Municipality of Boden and the National Library. The project resulted in a national centre for long-term digital preservation.

The delivery was divided into and discussed in three areas:

- transfer of the archival description;
- transfer of cases with documents compliant with eDoc; and
- selection and transfer of system and user documentation.

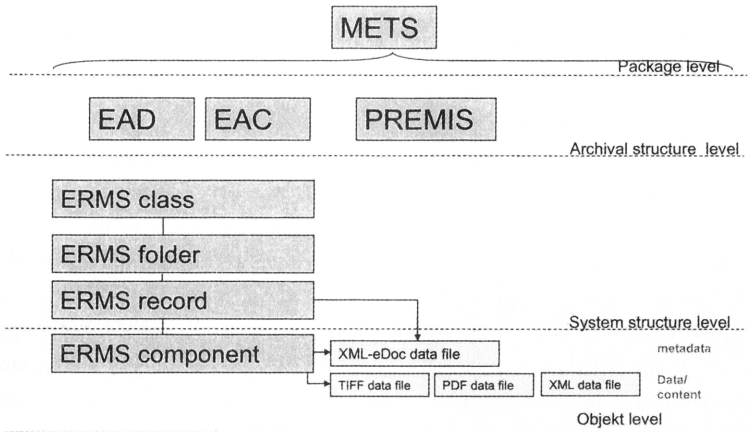

Figure 6. Schematic description of an SIP from the SIA to the National Archives

The trial delivery encompassed about 100 cases from various case management processes and was completed by the beginning of 2007. The delivery was packaged on the basis of OAIS. The entire trial delivery was packaged in a single SIP, thus differing from transfers within the SIA from the CMS to the long-term archive, where each case was viewed as an integral SIP.

In order to describe the SIP from the SIA to the National Archives, portions of the metadata encoding and transmission standard (METS)[15] were used. The METS file defined and described the files included in the package. The function of the METS file was considered to correspond to the package information (PI) and the package description (PD) of OAIS. For technical documentation of the various file formats included in the package, the National Archives used Preservation Metadata Implementation Strategies (PREMIS)[16] as its starting point. There was to be a PREMIS file for each file format or, alternatively, links to a register that would describe the file formats.

The transfer of the archival description was performed via EAD and EAC. The SIA's choice of the strategy to use ISAD (G) and ISAAR (CPF) made this part of the transfer relatively unproblematic. Unlike the SIA's archival description in the long-term archive, the National Archives' archival description system (ARKIS II)[17] contains special

authority records. This, however, did not cause any major problems. Information from the authority records was selected and delivered with an EAC file.

For the transfer to the National Archives, cases with appurtenant records were structured according to United Kingdom's National Archives electronic records management system (ERMS)[18] standard. The *ERMS class* level corresponds to the areas of activity and the case management processes, respectively, at the SIA. Information from these levels existed thus in both the ERMS files and the EAD/EAC files. The *ERMS folder* level described the individual case and the *ERMS record* described the documents in the case. Finally, the *ERMS component* described the actual data files in the package.

In the trial delivery from the SIA to the National Archives, the archival description was delivered from fonds down to sub-series in EAD and EAC. The case file level and the item or document level (according to the SIA's archival description) were delivered in ERMS files and thus not in EAD/EAC. The reason for this was that the National Archives' archival description system did not report farther down than to series level with regard to electronic documents. For the trial delivery, however, the National Archives produced a search interface that was very roughly similar to the SIA's CMS. Via this search interface, it is possible to refine a search in the agency archival description system, then to search directly, and view delivered cases and electronic documents that have been delivered to the National Archives through the ERMS files.

The selection of system and user documentation from the SIA's CMS and long-term archive for transfer to the National Archives gave rise to considerable discussion within the LDB project. The National Archives' requirements were significantly more rigorous than the SIA could meet with its delivery. There were several reasons for this. Among other things, the SIA lacked clear, efficient version management of system documentation. There were also difficulties in linking system documentation to a specific version of the CMS, which was under continuous, extensive development. A question of principle was raised during the collaboration between SIA, the National Archives and the LDB project, concerning the importance of voluminous, detailed system documentation. Perhaps it would be better to provide a more general

description of the CMS and focus instead on the information that was transferred through the cases. The transfer of the system documentation that could be found was performed in the traditional way by regular mail or email with no electronic link with the SIP from the SIA to the National Archives. The SIA and the National Archives agreed that much remained to be solved in the matter of documentation.

Summary and conclusions

During the period 2001–2006, a project for digital preservation was carried out in Sweden, encompassing the entire life cycle of information management. The case managers of the SIA had to handle most of the cases digitally when implemented in the case management system (CMS). When these cases had been closed, they were archived in a digital archive created by the SIA. Then, according to instructions and rules that had been produced, a sample of all cases was delivered to the Swedish National Archives for long-term digital preservation. A user-friendly register maintenance program was produced for establishing rules for destruction and preservation, and for archival description within the SIA project. There are a number of lessons to be learned from this project:

1. The use of standards is essential to all information management. Without standards, the wheel must constantly be re-invented, and the transfer of information from one authority or organisation to another will suffer.

2. Digital preservation requires active administration and monitoring of the surrounding world, since conditions change relatively quickly. Even if the file formats chosen by the SIA may be viewed today as quite stable, migrations and conversions will still be needed and, with them, active administration.

3. The archival description that has been produced with the support of ISAD (G) and ISAAR (CPF) will need to be modified on the basis of ISDF,[19] even considering the fact that a mere five years have passed since the project at the SIA was initiated. The use of these ICA standards was largely due to implicit demands from the National Archives, and it surely facilitated the transformation of the archival description in the

test delivery. To set this in perspective, a question must also be raised about also using instead the ISO standards developed during the time of the project (primarily ISO 23081 and ISO/TR 26122), for the archival description.

4. Even though half of the SIA's records are digital, a significant amount of paper records remain. Digital information management that is sustainable in the long-term entails great expense and it is not profitable today to digitise everything completely, optimism for the future notwithstanding. Therefore an archival report must be able to handle both digital files and physical volumes of paper records. This part needs to be taken into account much more than was done in the described project.

5. There are always continuing struggles about money between different interests in an organisation. Those responsible for long-term archiving matters therefore have an ongoing challenge to improve the arguments for why they are doing what they do for the organisation. Even if there is a great respect for laws and regulations, the risk of failure is big if economic gains for the organisation are not well-presented and effectively argued. This is a factor that records managers and archivists must take seriously if they are to have a chance at being successful in their work.

6. Although building and developing a digital archive may seem more fun than administering it, it is absolutely crucial that digital archives, including format handling, have an active administration and preservation plan. Without sound continuing administration and vigilant guardianship the archive will be placed at risk.

Endnotes

[1] The name 'The Swedish Insurance Agency' is used synonymously with the name 'The Swedish National Insurance Board' in this article.

[2] In paragraph 3 of the *Archives Act* (SFS 1990:782).

[3] The Secrecy Law (SFS 1980:100).

[4] Act (SFS 2003:763) on the treatment of personal information within the administration of social insurance.

[5] ibid.

[6] International Council on Archives, Committee of Best Practices and Standards, *ISAD (G): General International Standard Archival Description*, second edition, 1999, available at <*http://www.ica.org/en/node/30000*>, accessed 20 November 2009.

[7] International Council on Archives, Committee of Best Practices and Standards, *ISAAR (CPF): International Standard Archival Authority Record for Corporate Bodies, Persons, and Families*, second edition, 2004, available at <*http://www.ica.org/en/node/30230*>, accessed 20 November 2009. At the time that the archival description was produced, the International Standard for Describing Functions of Corporate Bodies – (ISDF) – was not yet in the picture.

[8] The formats we advocated were XML, PDF/A and raster graphics.

[9] *Reference Model for an Open Archival Information System*, CCSDS 650.0-B-1, 2002, available at <*http://public.ccsds.org/publications/archive/650x0b1.pdf*>, accessed 20 November 2009.

[10] *EAD: Encoded Archival Description*, 2002, available at <*http://www.loc.gov/ead/*>, accessed 20 November 2009.

[11] *Reference Model for an Open Archival Information System*, p. 30.

[12] Dublin Core Metadata Initiative, available at <*http://dublincore.org/*>, accessed 20 November 2009.

[13] *EAC: Encoded Archival Context: Corporate Bodies, Persons, and Families*, available at <*http://eac.staatsbibliothek-berlin.de/*>, accessed 20 November 2009.

[14] LDB: Centre for Long-term Digital Preservation and Access, available at <*http://www.ltu.se/ies/ldb-centrum?l=en*>, accessed 20 November 2009.

[15] *METS: Metadata Encoding and Transmission Standard*, available at <*http://www.loc.gov/standards/mets/*>, accessed 20 November 2009.

[16] *PREMIS: Preservation Metadata Maintenance Activity*, available at <*http://www.loc.gov/standards/premis/*>, accessed 20 November 2009.

[17] *ARKIS II: A Swedish Archival Information System*, available at <*http://www.knaw.nl/ecpa/sepia/workinggroups/wp5/1.html*>, accessed 20 November 2009.

ARKIS II was implemented 2003 and is still in use. It was preceded by ARKIS, which was implemented in the early 1990s. ARKIS II is used for registration and retrieval of archives delivered to the National Archives and also for making their own older collections more accessible. Improvements in the current version include a separate module for authority records and a more distinctive multi-level approach.

[18] United Kingdom's National Archives electronic records management system standard, available at <*http://www.nationalarchives.gov.uk/electronicrecords/reqs2002*>, accessed 20 November 2009. See especially requirements for ERMS, part 2, metadata standard.

[19] International Council on Archives, Committee of Best Practices and Standards, *ISDF: International Standard for Describing Functions*, 2008, available at <*http://www.ica.org/en/node/38665*>, accessed 20 November 2009.