

Swedish National Data Services (SND), the OAIS reference model and archaeological data

Ulf JAKOBSSON

Swedish National Data Service (SND)

Abstract: The Open Archival Information System (OAIS) model and its workflow provide the basis for SND's activities and how we handle research data. This goes for all kinds of data that are deposited at SND, including archaeological data.

The workflow comprises all activities from the moment when researchers deposit data at SND and the following quality control (Ingest); the enhancement with metadata, production of codebooks (Data Management); policies and recommendations (Preservation Planning); archiving the data for long time storage (Archival Storage); up to the moment when we make the data searchable via web interfaces and accessible for new research (Access).

The OAIS model has been applied, for example, for the data of more than 360 archaeological surveys (388 datasets; shape-files, reports, Access databases, >40,000 files, >7.3 GB¹), a few thematic databases of archaeological data, and it is used for long-time storage of parts of the Swedish Rock Art Archive material (>120,000 images, >5.8TB²). The survey data is from one (out of 21) counties and from the 21st century only. SND have only received archaeological data during the past two years and expect a tenfold increase within the next few years.

Since SND did not have to take care for archaeological data until 2011, questions arose about how it was structured, what metadata was needed (for SND as well as for the archaeological community), and how to document them had to be learnt and implemented. Furthermore SND have to inform the depositors about how they in turn could improve their data and metadata.

The GIS data are the first archaeological data that SND has made directly downloadable without any registration and it is free for anyone who wants to use them. These data are also among the first at SND where versioning has been fully implemented as well as geo-spatial information documented using the Data Documentation Initiative (DDI) standard.

Keywords: archaeology, archive, workflow, data, metadata

¹ Number per 2014-01-13

² Number per 2014-01-13

Background and organization of SND

The Swedish National Data Service was established 1st January 2008 by the Swedish Research Council. The organization is managed by the University of Gothenburg and is financed by the Swedish Research Council and the University of Gothenburg. SND is a national resource for the coordination of existing and newly established databases within the social sciences, humanities and medicine. SND is a digital archive based on the OAIS model.

SND originates in SSD (Swedish Social Science Data Service), a project that started in 1980 at the department of Social Science at University of Gothenburg and has thus more than 30 years of experience in documenting, quality assuring, disseminating and updating digitized research data. SND have today a staff of 31 members of which 2 are mainly working with archaeological data. SND started archiving archaeological data in 2011, but it is a rapidly growing area of work.

Open Archival Information System (OAIS) model

The OAIS and its workflow model provide the basis for SND's activities and how we handle research data. The model describes the workflow from ingest to dissemination and archival.

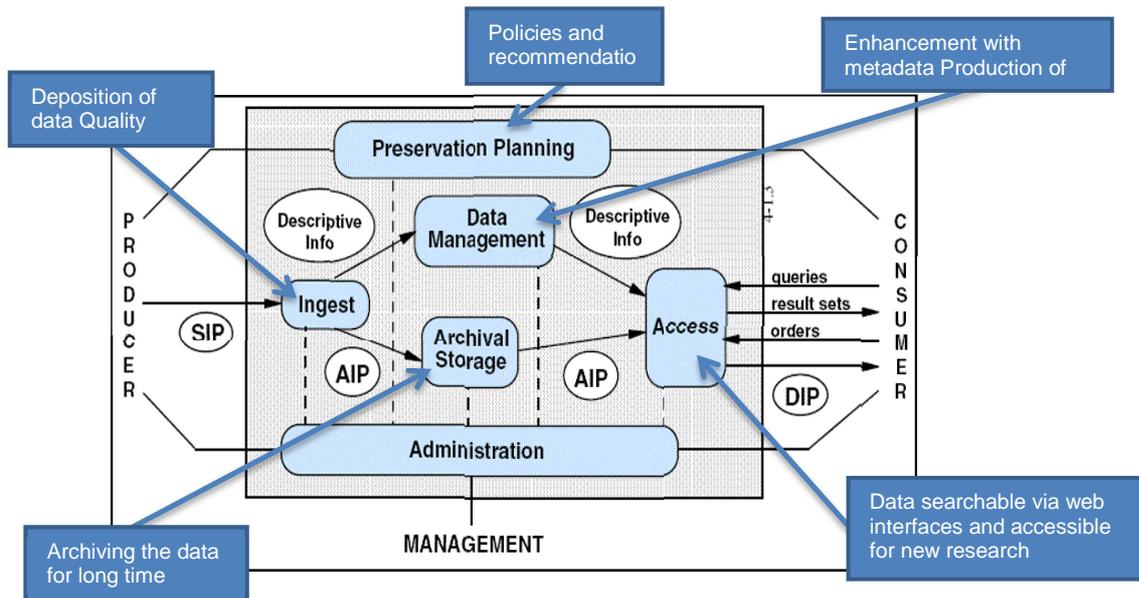


Fig. 1 – The OAIS model and archaeological data at SND (SND). Submission Information Package (SIP): package delivered by the Producer to the archive. Archival Information Package (AIP): package within an archive Dissemination Information Package (DIP): package sent by archives to the Consumer Descriptive Information: information that supports the finding, ordering, and retrieving of packages

SND, OAIS and archaeological data

The OAIS model has been applied for all data deposited at SND including the data to more than 360 archaeological survey and SND will continue to work according to this model. The survey data is from one

(out of 21) counties and from the 21st century only, but there is more data going to be deposited in the near future.

Archaeological GIS data and SND

Until August 2011 SND did not have any archaeological GIS data. There was no knowledge of how the data was structured, what kind of documentation or metadata we could expect or what kind of requirements we had to meet in order to securely store and disseminate the data in accordance with good practice.

However in August 2011 this changed. SND was contacted by a researcher at the University of Uppsala who was interested in depositing a vast amount of archaeological GIS data that was part of the project "Action program for the contract archaeology" initiated by the County Administrative Board³ in Östergötland, eastern Sweden (one of 21 boards).

The project, managed by the researcher Daniel Löwenborg, aimed to harmonize available GIS data so that it had, among others, a uniform structure, was located in the same coordinate system and that the files was converted from formats like Intrasis, shp, AutoCAD dxf into shp only.

Experiences from preparation of data/metadata

With the new type of material coming in there were several questions that had to be answered before the data was received: What kind of software was needed? How was the material documented? What information in the files could be used to describe and to further document the data? What information should be used as search criteria in the web catalogue and what information could be extracted directly from the files?

SNDs Internal Management System (SIMS) was built for and adapted to social science data, specializing in surveys. Several major changes and additions to the internal management system, but also SNDs web catalogue, was needed: adaptation to the concept "Geographical information" like coordinates, bounding box, polygons; administrative areas in hierarchies (church parish, civil parish, provinces, counties, municipalities); map search, time periods, new system of keywords (previously only ELSST, European Language Social Science Thesaurus, used by CESSDA, the Council of European Social Science Data Archives) with archaeological terminologies as well as the possibility to directly download data from SNDs web catalogue without registration.

³ The County admin. boards are giving permissions to archaeological excavations and surveys. The boards are also the controle authority for how the archaeological material is handled.

In March 2012 SND received the first package of data (shape-files, access databases and reports). It turned out to be merged GIS data to 230 rescue/contract archaeological surveys. The data to these surveys originated from 5 different organizations (private archaeology companies, museums and different sections of the National Heritage Board), covering 269 659 square meters, 12 162 archaeological features. After discussions the project manager re-deposited the material two months later, this time as individual files (not merged). The first 20 studies were published in our web catalogue in early June.

Required preparation of data/metadata by the archaeologists for ingest

While working with the material for these first surveys we noticed several things: the input of values differed depending on original organization and on the individual operator. We discussed this with the project manager and agreed that several major changes had to be done to increase machine readability (easier/faster publication). The two most important changes as SND saw it was how the date was set and how geographical information was written.

TID	T
2000	AU 2/FI
2004-04	AU etaj
2004-11 - 2005 - 04	FU
2004-11	AU etaj
2004-05	FU
2004-10	AU etaj
2004-11	FU
2004-09	AU etaj
2004-08	AU etaj
2004-11	FU del
2004-10/2005-04	AU etaj
2004-05/06	FU
2004-09 - 2004-10	AU etaj
2004-10/2005-06	Forsk.u
2005-06, 2007-07, 2007-08	FU
2005-03	FU

Fig. 2 – How date was set (SND)

PLATS
Östergötland, Slaka, Vikingstad, Viby, Sya, Mjölby
Östergötland, Motala-Mjölby, Russingstorp omr 8
Östergötland, Slaka, Lambohov 2:20
Östergötland, Linköping, Tannefors 1:107
Östergötland, Ekeby, Trimplaberget 2:40 mfl
Östergötland, Vårdberg, Tannefors 1:107
Östergötland, Vårdberg, Tannefors 1:107
Östergötland, Linköping, Tannefors 1:107
Östergötland, Linköping, Tannefors 1:107
Östergötland, Slaka, Lambohov 2:20
Östergötland, Slaka, Lambohov 2:20
Östergötland, Landeryd, Vårdsberg, Tannefors 1:107
Östergötland, Fivelstad,
Östergötland, Västra Stenby, Landstorp

Fig. 3 – Geographical information (SND)

How to set the date for the excavation/survey

The date was set in several ways in the files:

- Spring 2001
- May 2000
- April-June 2006
- May 1 to 15 2002
- 2001-08 - 2001-11
- 2003-07, 2003-11
- 2004-10/2005-04
- 2004-05/06

where comma (,) or slash (/) meant “and”, hyphen (-) for “from... ..until” and so on. It varied a lot.

In our opinion the date should be given in ISO format like: 2001-05 = the excavation/survey occurred in May

2001-05/2001-07 = the excavation/survey occurred from May till July (period) 2001-05, 2001-07 = the excavation/survey occurred in May and in July (can also be written with a colon) Even better if the start/end date is given in different fields.

Geographical information

The geographical information in the data was given under Plats (place). Here one could find information about counties, provinces, municipalities, civil parish and at what real property the survey/excavation took place. However the information was incoherent. In most cases information about municipality was missing, the information about county/province was given as one (even though it can differ) and since some municipalities have the same name as the civil parish it was not possible to read the text as comma-separated information. Initially the input of this information into SNDs management system was done manually.

Since this information could be used as search criteria in SND's web catalogue the information had to be divided into several labeled fields. It was also important to make it possible to add this kind of information to SNDs management system. After discussions with the project manager we came to the suggestion that four new fields should be added to the data: County, Province, Municipality and Civil parish. Information about real property would remain in the existing field.

The first three lines in Fig. 3 would then be:

County	Province	Municipality	Civil parish
Östergötland	Östergötland	Linköping, Mjölby	Slaka, Vikingstad, Viby, Sya, Mjölby
Östergötland	Östergötland	Motala, Mjölby	Russingstorp
Östergötland	Östergötland	Linköping	Slaka

Tab. 1

Another thing we started to look into was the possibility to present the survey/excavation area on a map as well as some tool for map search functionality. Since the system we had for documenting data could not (at that time) present information via polygons we decided to use a system with Bounding Box so that it at least was possible to get an idea about where the survey/excavation had been conducted. For this we needed information about x-min/max, y-min/max for the excavation/survey area. The information existed in the files but was not visible in any field. Agreement with the project manager was made to add coordinate fields for x-min/max, y-min/max and to visualize the information in the GIS-files.

During the autumn of 2012 these coordinate fields was added to the GIS-files so that the, in the management system, newly implemented Bounding Box system could be used. In the deposited material, the format for date was changed. SND and the project manager further agreed that the place information should be divided into the suggested 5 fields. The manual work of dividing the information into different fields in the GIS-files begun. Data to another 157 new studies was deposited to SND. In this new material coordinate fields to be used for the Bounding Box system was added to the files by the project team. The

first 20 studies was re-published (new versions with divided place information and added Bounding Box information) in August as well as 200 new ones.

In October 2012 SND received data to another 110 surveys/excavations that were published in the same month. Data to another 47 surveys/excavations was deposited in December. In this data the geographical information was separated into 5 fields and the information to Bounding Box was added.

In January 2013 was a hierarchical system for the geographical information (all counties, provinces, municipalities, civil parishes and church parishes in Sweden) implemented in the SND management system which prepared the possibility to later on ingest information from the GIS-files when that work was done. In order to enable citation of data, a system for assigning persistent links to studies and datasets was created in December 2013. Data can since then be linked to other information such as related reports, papers or other relevant data. The “final” 47 studies were published.

Access and re-use of archeological data

It was important for the project manager that the GIS data were directly downloadable without any registration or contact for anyone who wanted to use the material. These functions had to be implemented to our systems. Another important feature that was introduced and implemented during the process was versioning of the data.

Most of the information in the web catalogue is extracted from the shape files and the reports. The information can be seen (and searched for) in SNDs web catalogue⁴.

There is also information about the dataset which is a bundle of several shape files together with an access database. The access database contains attribute data and information about the relationship between objects. The datasets and the excavation reports are directly downloadable from the web catalogue. Even though the data file and the reports and in Swedish the metadata is partially translated to English.

The material will be searchable via the ARIADNE infrastructure.

⁴ <http://snd.gu.se/en/catalogue/search/archaeological>

Find and order data

- Order data >
- Question bank (beta) >
- Open Data
- SND Nesstar Server
- International data
- Swedish party programmes and election manifestos >
- The Swedish Gallup archive 1942-1954 >
- Archaeological GIS Data
- Data of the Month >
- Accessibility levels

GIS-material for the archaeological project: Graves and burial grounds at Linköping Airport

Description
Documentation
Publications
Related studies

Study number: [SND 2004](#)

DOI: [doi:10.5878/000511](https://doi.org/10.5878/000511)

Title: GIS-material for the archaeological project: Graves and burial grounds at Linköping Airport

Principal investigator(s): [Swedish National Heritage Board, UV Öst](#)

Purpose:
 The information in the purpose is translated from the archaeological report:

 The preliminary investigations aimed at clarifying the nature and the burial customs of the recorded graves and burial grounds, as well as any new-found relics within the areas. The aim was also to demarcate the respective archaeological area and to assess the extent of the remains in them. The results will also be used as a basis for assessing subsequent outcome study needs.

Abstract:
 The information in the abstract is translated from the archaeological report:

 The Swedish National Heritage Board's contract archaeology division UV Öst has, during the spring of 2001, conducted preliminary archaeological surveys of a number of ancient monuments in the area of Linköping Airport. The investigations were prompted by the upcoming development in connection with the take-off and landing runway. The ancient remains consisted of single graves and groups of graves, located on the remaining heights that currently surround the airport. The investigations showed that the three sites were burial grounds, while one site consisted of a smaller group of graves. In essence, the graves are dated to the Early Iron Age. Several of the sites also contained findings from the Mesolithic.

Topic classification: [Archaeology, History and Archaeology](#)

Keywords: [Bronze Age](#), [Early Iron Age](#), [Iron Age](#), [Migration Period](#), [burial ground](#)

Mode of data collection:
 2001-04 (Field)
 2001-04 (Field)
 2001-04 (Field)
 2001-04 (Field)
 2001-04 (Field)

Map:



Leaflet | Map data and imagery © OpenStreetMap contributors, CC-BY-SA, [OpenStreetMap](#)

Kind of data: GIS

Metadata: [DDI Lifecycle 3.1](#), [DDI Codebook 2.1](#), [MARCXML](#)

Availability status: 1a - Freely available without registration

Fig. 4 – Catalogue post, SND (SND)

Datasets

001 **Saab 125**

Version: 1.0, 2012-09-12 [doi:10.5878/001566](https://doi.org/10.5878/001566)

Description: The ZIP file consist of GIS files and an Access database with information about the excavations, findings and other metadata about the archaeological survey.

Kind of data: GIS

Languages:

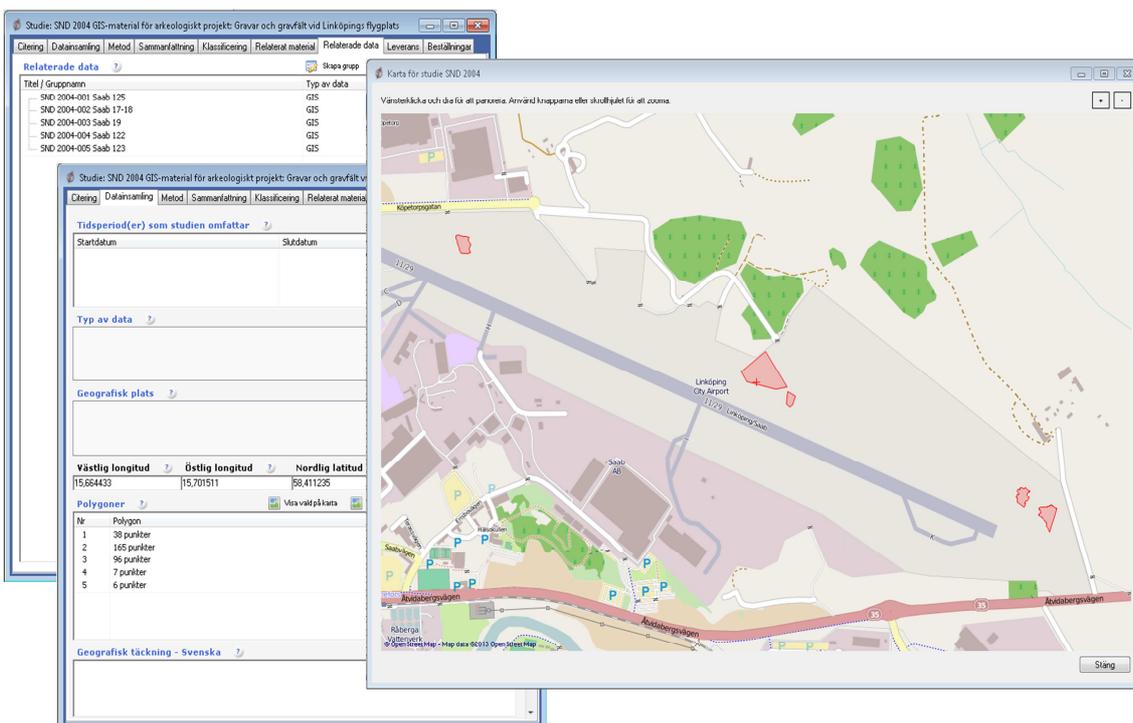
Languages	
data	Swedish

Data:  [SND 2004-001-V1_0.zip](#)
Saab 125

Fig. 5 – Dataset information, SND (SND)

Maps

As mentioned earlier, SND uses the documentation standard DDI to document and describe the material and the lifecycle of the data. However it is also possible to use this standard for geographical information and, for example, to produce polygon information on maps. It is possible to extract polygon information from the shape-files into a XML-schema and DDI and then produce that information in a map in an open source system.



The screenshot shows a GIS application window titled "Studie: SND 2004 GIS-material för arkeologiskt projekt: Gravar och gravfält vid Linköpings flygfält". The interface includes several panels:

- Relaterade data:** A list of data items with columns for "Titel / Gruppnamn" and "Typ av data".

Titel / Gruppnamn	Typ av data
SND 2004-001 Saab 125	GIS
SND 2004-002 Saab 17-18	GIS
SND 2004-003 Saab 19	GIS
SND 2004-004 Saab 122	GIS
SND 2004-005 Saab 123	GIS
- Tidsperiod(er) som studien omfattar:** Fields for "Startdatum" and "Slutdatum".
- Geografisk plats:** Fields for "Västlig longitud", "Östlig longitud", and "Nordlig latitud".

Västlig longitud	Östlig longitud	Nordlig latitud
15.664433	15.701511	58.411235
- Polygoner:** A list of polygons with columns for "Nr" and "Polygon".

Nr	Polygon
1	38 punkter
2	165 punkter
3	96 punkter
4	7 punkter
5	6 punkter
- Geografisk täckning - Svenska:** A dropdown menu.

The main map area shows a geographical view of the site with various colored polygons overlaid on a street map. Labels on the map include "Linköping City Airport", "Linköping Flygfält", "Saab 125", "Råberga", "Åsöbergsvägen", and "Åsöbergsvägen".

Fig. 6 – Catalogue post, SND (Copyright: SND)

Conclusion: In the future

Planned development for the GIS material is to expand the search options through, for archaeology/history, a better adapted filtered search system allowing different ways to narrow and expand the search. The plan is also to implement a map search, where you for example can select an area on the map and find out which excavations was made in this particular area. SND also examines other possibilities to search by map including a combination with filtered search, and allows other search services to harvest SND entries so that the material is visible in other relevant forums.

Other plans for SND is the development of a new management system based on modules. There are several reasons for this. One is that since SNDs management system is used for data management, documentation, administration etc. i.e. anything that one has to do at the archive, the software gets more and more advanced. Depending on what work has to be done different part of the software is needed. The idea is that only those parts that are needed starts and by that decrease the workload on the software and by that minimize the waiting time that might occur. The software today was initially created for the management of social science data (mostly surveys) and has been adapted over and over again for new kinds of data.

SND is also looking into better and more flexible system for data deposition, an automated system for assigning persistent identifiers.

References

SNDs website: <http://snd.gu.se/en>

Archaeological data in SNDs web catalogue: <http://snd.gu.se/en/catalogue/search/archaeological>

OAIS reference model, ISO 14721:2012: http://www.iso.org/iso/iso_catalogue/catalogue_ics/catalogue_detail_ics.htm?csnumber=57284

ELSST-catalogue: <http://elsst.esds.ac.uk/>

CESSDA portal: <http://www.cessda.net/>

ARIADNE infrastructure: <http://www.ariadne-infrastructure.eu/>

Imprint:

Proceedings of the 18th International Conference on Cultural Heritage and New Technologies 2013 (CHNT 18, 2013)

Vienna 2014

<http://www.chnt.at/proceedings-chnt-18/>

ISBN 978-3-200-03676-5

Editor/Publisher: Museen der Stadt Wien – Stadtarchäologie

Editorial Team: Wolfgang Börner, Susanne Uhlirz

The editor's office is not responsible for the linguistic correctness of the manuscripts.

Authors are responsible for the contents and copyrights of the illustrations/photographs.