

Elisabeth Verhelst\*, Liesbeth Missel\*, Bas Vanmeulebrouk\*\*, Frans. I. Rip\*\*\*

## **Publishing WWII aerial photographs in geographical and library information systems**

*Keywords:* WWII; aerial photography; geo portal; metadata.

### *Summary*

The Library of the Dutch Wageningen University and Research centre houses a collection of aerial photographs taken by the Allied Air Forces. The collection is part of a project that aims to publish these images in a user friendly way so that they are accessible to a wide audience. This paper describes the nature of the collection and its particular requirements for publication, as well as the design and development of the application for collection presentation. The collection's geographic characteristics dictated that the collection should be presented with other geographic materials such as digital historical maps and actual aerial photos. To create this information compilation, technologies from geographical information management were combined with technologies from library automation. This involved mainly metadata specification and management as well as usability design.

### **The WWII RAF aerial photograph collection**

Since 1994 the Library of Wageningen University and Research (UR) centre houses a collection of aerial photographs taken by the Allied Air Forces, popularly known as RAF-collection<sup>1</sup>. The date of the photographs taken lies mainly between 1944 and 1945. The collection came in our possession when they no longer were of use for the Dutch institute of soil mapping in Wageningen, who received the collection after World War II with the intention to use them as source information to produce soil maps. Over the years several parties had shown their interest in these images, amongst them the Dutch Explosive Ordnance Disposal Service (EOD) and the National Archive. Because of the relevance of the pictures for research on land use the collection was donated to the library of the Agricultural University, which became later our current library. The collection consists of 94.257 photos, taken with different cameras, on different heights and different angles, mainly vertical but some are taken oblique. Some of the photos are distorted due to hazardous conditions during the reconnaissance flights and the hasty photographic production and use. The photos were taken by RAF pilots who flew in a Mosquito airplane in *sorties* (flights) of several runs of tens of photos. Within a run the photos overlap 60 % so that they can offer 3D images when studied with a stereoscopic viewer. The geographic position of the sorties is also documented on sortie maps.

---

\* Wageningen UR Library, the Netherlands [lieke.verhelst@wur.nl]

\*\* Alterra Centre for Geo-Information, Wageningen, The Netherlands

\*\*\* Wageningen University, Wageningen, the Netherlands

<sup>1</sup> <http://library.wur.nl/speccol/aer-phot.html>

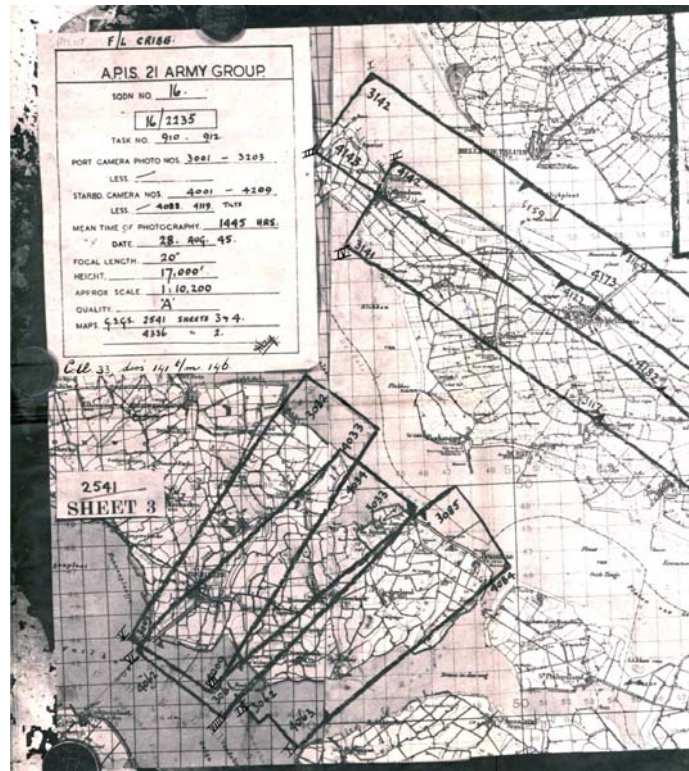


Figure 1. Example of a sortie map. The location is Zeeland, province of The Netherlands.



Figure 2. Example of an aerial photograph. Location is near Apeldoorn, The Netherlands.

As the ideas for a digital presentation to the public became more realistic, a first step that we took was the digitizing of the analogue photos to a resolution of 1200 dpi and serving them on a plain website in 150 dpi samples. This was done with a granted subsidy within the War Heritage Programme (Erfgoed van de Oorlog) from the Ministry of Health, Welfare and Sport because of the collection's large cultural and historical value. The purpose of this website was merely to create visibility of the collection and to set-up a web-shop to lower the need for handling the original photographs. Once this task was completed we had a digital collection storage of 11 TB of data. But the way to search the collection for relevant photographs still had to be done by using the old digitized sortie maps. This originated the idea to create a geo portal according to the wishes of our clients: to be able to locate the photographs geographically and to integrate the collection with other geographical material like actual topographic maps and landscape plans. We also wanted to offer a restricted view on the high resolution scans so the use of the vulnerable originals could be made redundant almost completely.

### **Development of the discovery application**

As the objective of the project was to create a digital discovery environment that could serve the interest of a variety of users, the technical requirements of the collection application had to satisfy a variety of interest groups. During the design phases in the project we kept this objective in mind. Phases in which important decisions were made were the georeferencing of the collection, the metadata management design and the specification and design of the user interface. The latter is still work in progress.

The architecture of the collection discovery application is based on three layers: the actual data (photographs and metadata), the data and metadata servicing layer and the presentation layer (user interface). The application can be used in a Spatial Data Infrastructure (SDI) and a Library Information System (LIS).

#### *The data layer: positioning the photographs on the map*

In order to project the aerial photographs on a digital map the material had to be georeferenced. Normally this is a manual exercise performed by specialised staff. In this process an image is manually positioned, using a GIS tool, upon a reference map. Geographic coordinates are assigned by clicking reference points on the image and link them with corresponding points on the reference map. Depending on the distortion of the image, the projection of the reference map and the accuracy requirements the number of referencing points is at least two. Because of the enormous number of photographs to be processed we aimed for the design of a methodology that was more cost effective than manual georeferencing. The eventual method consisted of still a substantial portion of manual work, such as checking the quality for processing of the almost 95 thousand images as well as bits of manual georeferencing, and automated work. We developed a script that performed the definitive georeferencing and image positioning.

The input files for the total process were the maps showing the geography of the flights ('sorties'). The start- and endpoint photographs of those sorties were georeferenced by hand. We used these as reference to automatically calculate the position of every photograph taken between the begin and end point of each flight. Some flights appeared to be deviating from a straight line, probably caused by wind or situational military dangers (Fig. 3). This brought about the necessity

to manually georeference some more intermediate points positioned between the begin and end points.

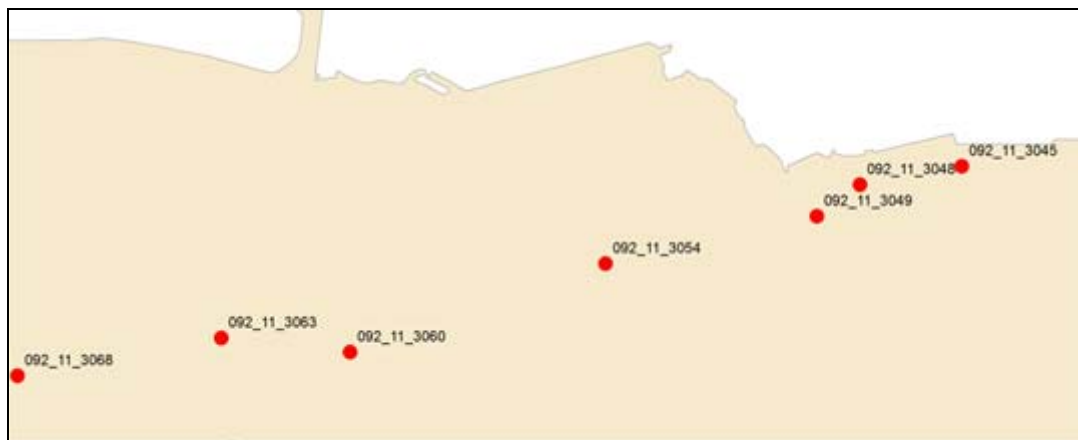


Figure 3. Off track position of images of one flight (number 092\_11) causing the necessity of manually georeferencing intermediate points.

The begin and endpoints were input for an automated script that calculated the position of every photograph in between. The calculated position was used to automatically georeference the photograph.

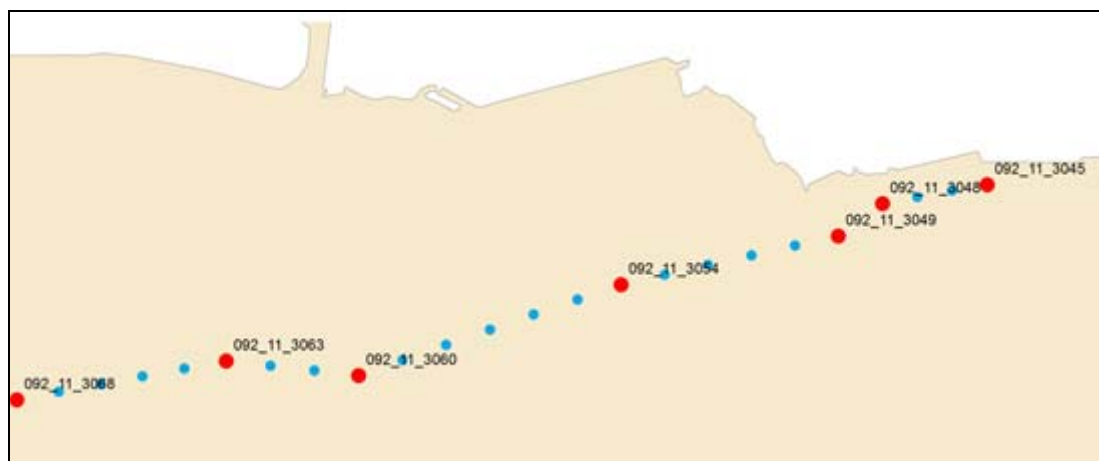


Figure 4. Automatic positioned images (smaller dots) added to manually georeferenced images (larger dots).

A second script took care of the rotation of the images, necessary to make the pictures fit in within the cartographical background. The calculation of the rotation was derived from the direction of the flight. Both scripts were developed with the Python programming language in combination with ArcGIS geoprocessing tools.

#### *The data layer: metadata*

For discovery of the collection we wanted not only to offer a visual geographic interface – the map of the Netherlands - but also textual retrieve options based on metadata stored with individual photographs. Users search with a certain goal in mind and within their own context (Rose and Levinson 2004). It is evident that the metadata elements determine which information about the accompanying dataset one is able to retrieve, so the composition of a metadata element

set must be in accordance with the search behaviour of a user group or harvesting system. This is particularly valid for geo information objects, to which the RAF aerial photographs belong, because they can be of interest to a large and varying audience.

Thus, a metadata element set was developed with these requirements in mind. The most important step in this process was the definition of the end user groups. To each end user group a collection of metadata elements was assigned. In most cases we were able to use a standardised metadata element set. For some we had to define additional elements. An overview of the end user group and the corresponding metadata element set we assigned is given in Table 1.

<b>End User Group</b>	<b>Metadata element set</b>
Cultural heritage (Europeana)	Dublin Core and the Europeana Semantic Elements
Geo Information Scientists	ISO19115 and ISO19119 standards (definition) and ISO19139 (XML format)
General public	ISBD(CM) International Standard Book Description (Cartographic Material)

Table 1. End User groups and related metadata element sets.

For publishing the metadata in Europeana we used the Dublin Core set as part of the Europeana Semantic Element Specification ESE (EU 2011) and the Europeana Data Model (EDM) elements, (version 5.2.3, 24/02/2012). Europeana's aim is to accumulate digitised content, standardise the data that describes it, apply linked data techniques to enrich it, and promote persistent identifiers to locate it in the long-term (Europeana 2012). Following their published Content Strategy, they extended their network of data providers and encouraged the development of aggregators that fit the needs of different countries, domains and users. In the Netherlands our RAF collection can be aggregated by either the War Heritage Sources or the Dutch State Institute for Cultural Heritage or maybe both.

To enable Geo Information Scientists to use and exchange services for the aerial photos in their spatial data infrastructures (SDI's) the aerial photos must be opened up in a interoperable fashion. Interoperability refers to the ability of one system to interact with other systems in a reliable and unambiguous manner. Within the spatial domain, the open standards of the Open Geospatial Consortium (OGC) are widely used. This is in line with the European INSPIRE directive that regulates that common implementing rules are adopted by member states. It is for that reason that we used the OGC standards. Geographic metadata is served via the OGC metadata Catalogue Service for the Web (CSW) implementation specification. CSW's support the ability to publish and search collections of descriptive information (metadata) for data, services, and related information objects (Open Geospatial Consortium Inc. 2007). Different metadata implementation standards can be used with a CSW. In order to comply with the European INSPIRE regulations, in our project the ISO 19115 and ISO 19139 metadata implementation standards have been used. Lastly, the general public of the Wageningen UR Library must be able to find the photo's via the common user interface of the library<sup>2</sup>. The metadata on which the search facility for this collection is based is the ISBD(CM) metadata format.

The RAF photo collection itself was described with limited information. Less than 10 metadata elements were present describing the details of the flight, such as flight date, pilot name and

<sup>2</sup> <http://library.wur.nl>

sortie number. These elements were added to the total collection of user targeted metadata sets from Table 1.

The complete metadata element set consists of approximately 35 elements. The elements are organised in a nested structure in order to group them together and use them in the appropriate context levels. These levels are: general metadata elements which will apply to virtually any dataset, metadata elements which will apply to almost any dataset with a spatial reference, metadata elements which will apply to all raster datasets with a spatial reference and metadata elements which are specific to this collection.

To describe the metadata for the RAF aerial photos, an XML Schema Definition (XSD) was created. An XSD is used to describe the contents of XML files.

#### *Serving the image data*

For serving the aerial photographs we used the OGC Web Map Service (WMS) and Web Coverage Service (WCS) implementation specifications. A WMS dynamically produces maps of spatially referenced data, such as our georeferenced images. A "map" is in this case a portrayal of geographic information as a digital image file suitable for display on a computer screen (Open Geospatial Consortium Inc. 2004). The WMS standard is merely used for data visualization. A WCS on the other hand describes and delivers multidimensional coverage data over the internet. A coverage is a georeferenced raster, for instance gridded geospatial data, or a collection of remote sensing images or aerial photos (Open Geospatial Consortium Inc. 2008). A WCS provides access to the actual data.

#### *Serving the metadata*

In our application we decided not to present to the user the metadata set that contained every element. This would overload a user with irrelevant information. Instead we decided to store all the metadata information in one combining metadata dataset and present to the user a subset of all this metadata information based on the search context of the user.

This concept is supported by the general architecture of our LIS. This is based on a three tier environment of which the XML database is the first tier. The full metadata set is stored here. The second layer is a content management system based on WebQuery and merely regulates the store and retrieve requests to the XML database. WebQuery accepts either a URL or a form, containing fields that represent a query. The query will be passed to the database backend, which will perform the query and return a record-set. Records are always stored as XML records. WebQuery can be instructed to use XSLT (Extensible Style Language Transformation) files to transform the XML records into other formats. This architecture supports not only serving appropriate metadata formats to any user group, but it also supports serving any metadata format to a harvesting system.

The presentation layer is based on XSLT. XSLT files are used to transform the full RAF collection XML metadata documents to other formats (in our case HTML for presentation mark-up in the library user interface, Dublin Core XML for description of content for Europeana and ISO19139 XML for use in SDIs).

The metadata will be indexed for search in the library central indexer. This indexer, based on Apache Lucene SOLR<sup>3</sup> allows us to present to the user a ranked and faceted result set. This guarantees that the individual aerial photographs are also findable from the central library text search interface.

### *Presenting to the user*

The enormous amount of photos scattered over a large area of The Netherlands and sometimes positioned on top of each other, demands for a well-balanced functional design of the presentation of the collection. While this is still work in progress we are currently building a list of required functionalities. The final layout of the interface will be realised via discussions in a series of workshops to be held with representatives of user groups. The most important search tasks that we already identified are:

- searching for photographs of a specific location,
- searching for photographs via a description in the metadata,
- browsing through the runs, exploring areas and features of interest

To facilitate these search tasks, the interface will offer a search facility for text entries stored in the metadata as well as a geographical search facility based on a map of the Netherlands. The map search will support both interactive search (panning, zooming) and search via entering the location name in a map search interface.

Since the images are positioned on top of each other the total collection will not be shown as images in the geographical interface, but rather via all the image location points as illustrated in Figure 4 .

### *Conclusion and discussion*

The main challenges that we were facing at the start of this project were to combine the geospatial perspective with the library perspective. While we were working, we learned that their methods are the same, but the use of the content is very different. Technologies like metadata management and harvesting, web services and indexing are alike. Since the content is used differently, different metadata formats and presentation techniques are needed. In order to manage the metadata content in both infrastructure perspectives (SDI's and LIS), we have used a methodology for metadata management that serves both worlds.

Some users search via geographical interfaces. Some use text. We aim to develop an interface that satisfies both types of users. In addition we have plans to provide the users with facilities to submit features that they discover on the photographs into the metadata. We expect that this type of crowdsourcing feedback will enrich the collection as a whole and improve its discovery.

The collection of photographs is very large. This brought us to develop an automated method for georeferencing. Obviously, the automatic georeference calculation provides an idealized and average result. Our calculation is based on assumptions such as a constant flying speed, camera height and camera angle. In reality, this may not always have been the case. The georeferencing results are not as precise as they would have been when done manually. We found that zooming to a scale over 1:50:000 will show deviations. However, taken into account the cost versus the results, we have assumed that these results are sufficiently satisfying for the interest groups. To

---

<sup>3</sup> <http://lucene.apache.org/solr/>

verify this assumption we have plans to invite the users to indicate which areas they consider candidates for more precise georeferencing. These plans include the development of tools for user generated georeferencing.

### *Acknowledgement*

Quality verification and manual georeferencing of the collection was done by Gert van Dorland. The software for calculating intermediate image positions and their rotation was created by Rini Schuiling. Both are with the Alterra Centre of Geo-Information at Wageningen University and Research Centre.

### **References**

- EU (2011). Europeana Semantic Elements Specification. *Version 3.3.1* 24/01/2011.
- Europeana. (2012). "Provide data." 2012, from <http://pro.europeana.eu/web/guest/provide-data>.
- Rose, D. E. and D. Levinson (2004). Understanding user goals in Web search. *Thirteenth International World Wide Web Conference Proceedings*, WWW2004.
- Open Geospatial Consortium Inc. (2004). OpenGIS® Web Map Server Implementation Specification, from <http://www.opengeospatial.org/standards/wms>
- Open Geospatial Consortium Inc. (2007). OpenGIS® Catalogue Services Specification, from <http://www.opengeospatial.org/standards/cat>
- Open Geospatial Consortium Inc. (2008). Web Coverage Service (WCS) Implementation Standard, from <http://www.opengeospatial.org/standards/wcs>