NATO OPERATIONAL RECORD: COLLECTIVE ANALYTICAL EXPLOITATION TO INFORM OPERATIONAL ANALYSIS MODELS AND COMMON OPERATIONAL PLANNING FACTORS

1.0 SAS-100: BACKGROUND, OBJECTIVES, AND PROCESS

1.1 Background

The success of current and future NATO operations is positively influenced by operational analysis methods, models, and tools that rely on quantitative and qualitative data of operational records from past and current operations.[1] Therefore it is of critical importance to promote the collection of – and ready access to – operational records. This line of thought is implicit in NATO's Information Management (IM) policy that is in essence expressed in eight NATO council documents. Stressing that "records are critical to reliable assessment of operations both during their conduct and after their completion", *NATO Directive on the Management of Records Generated on Operational Deployment* (C-M(2012)0014) focuses on the collection, management, and archival of operational records by the strategic and operational commands.

NATO document C-M(2007)0118 (p.1) stresses that one of the goals of NATO's IM is "to support the effective and efficient use of information". In order to achieve this, information "must be organized in a standardized way that makes information easily discoverable and accessible, and must be managed as a corporate resource[2]". Moreover, information systems shall[3]:

- Ensure *easy access* to information respecting restrictions imposed for security or sensitivity reasons;
- Ensure the *timely availability* and dissemination of accurate information to users, organizations and systems;
- Enable use, re-use, fusion and *exchange of information* by both people and systems; and
- Allow for *effective and efficient discoverability* of relevant information.

NATO's IM policy requires that information is managed with an emphasis on the "responsibility-to-share" balanced by the security principle "need-to-know", all in accordance with security, legal and privacy obligations. Moreover, "the default position should be to share the information as widely as possible [...] unless there are compelling reasons" against this. In essence, information guarding and information sharing are regarded as mutually compatible, but there is also a priority ("default position") towards sharing.[4]

1.2 Objectives

SAS-100 Specialist Team adheres to the above policy in that a comprehensive, easily searchable, and readily accessible digital archive of operational records is regarded as presenting planners, analysts and other users with an opportunity for exploitation of past and current operational records. As mentioned in the SAS-100 Technical Activity Proposal (TAP), if followed, NATO's IM directive could foster a reality where strategic, operational,

[1] See NATO Directive on the Management of Records Generated on Operational Deployment (C-M(2012)0014).

[2] NATO document C-M(2008)0113, Annex, p. 1-9.

[3] Directly quoted from NATO document C-M(2008)0113, Annex, p. 1-11. Emphases added.

[4] See NATO document AC/322-N(2010)0026-REV1, Annex, p. 3.

and tactical records relating to an operation could be readily available and used across NATO and NATO Troop Contributing Nations (TCN). The result would be a unique, highly relevant, searchable and easily accessible, digital repository of qualitative and quantitative data/information detailing various NATO operations. Expressed differently, the repository would meet the IM directive's requirements of *easy access*, *timely availability*, *exchange of information and effective and efficient discoverability* of relevant information.

NATO's IM directives highlight three causally interrelated challenges for information sharing: rules, willingness and ability. The focus of SAS-100 is on ability, and is in that regard tasked to achieve two objectives that directly reflect the current NATO IM directives.

Objective 1:

- Establish a suite of procedures that allows for easy search, access, and immediate retrieval of archived documents across various military commands, non-military entities, and national governments through a computer platform independent solution;
- The procedures should include the identification of the types of archived data, the owner of the data (NATO or TCN), and the associated relevant level of detail of metadata;
- Where possible, the metadata should include releasability, format, quality of the data, time period, mission name, organization (military unit or non-military entity), and document type;
- The feasibility of and NATO demand for searchability in terms of detail of metadata and/or full text of the digital archive shall be explored;
- The solution shall not involve investment costs in terms of software for client computers and it shall meet NATO security requirements; and
- The solution shall be able to handle already archived documents, as well as documents from on-going missions that are not yet archived.

Objective 2:

- Identify the actors/organizations what would take a collective responsibility or oversight of the proposed digital archive, and its update; and
- Develop the means by which this responsible organization would make NATO and TCNs aware of the existence of non-NATO Archives and how to exploit them.

1.3 Process

The primary focus of the SAS-100 Specialist Team was on the tactical, field level records, with a secondary focus on strategic level records as they provide the larger context and background for tactical, field level events. Its work covered NATO records as well as NATO TCN records, for current and legacy operations. Digital repositories are common within civilian sectors and government agencies in some countries. This means not only that this report can draw on lessons from existing solutions, but also that the general approach of digital document access/retrieval is a tried, tested and increasingly utilized tool.

The work was carried out through three 2-day Workshops generously hosted in Paris (NATO STO/CSO, November 2012, February 2013) and the Hague (NATO Communications and Information Agency (NCI), May 2013), with expert participants from the United States, United Kingdom, France, Germany, Norway, Canada, Denmark, Netherlands and Sweden. The participants represented their countries and/or NATO agencies such as Supreme Headquarters Allied Powers Europe (SHAPE), International Security Assistance Force (ISAF), ISAF Joint Command (IJC), Joint Analysis and Lessons Learned Center (JALLC), NATO Archives, NATO

Communications and Information Agency (NCI), and Allied Rapid Reaction Corps (ARRC) in Brussels. The report is thus a multi-nation and multi-agency collaborative product, and has been coordinated by the Folke Bernadottte Academy, which is a Swedish government agency placed under Ministry for Foreign Affairs. It has been guided by the ambition to include a wide spectrum of insights from the legal, archival, analyst and field expertise communities. In this report discussions on technical issues have been reduced to improve readability and concreteness.

This brief and non-technical report is organized as follows. Section 2 discusses NATO records and IM management systems in general, whereas Section 3 offers a more detailed overview of not only NATO and US IM systems, but also some illustrations of civilian sector digital repositories. It should be stressed that the US IM systems as well as civilian solutions presented are illustrative rather than constituting an exhaustive review. In line with the objectives of SAS-100, Section 4 discusses an envisioned NATO digital repository with regard to contents, formats, and searchability; access, classification and security issues are discussed in Section 5; legal considerations are highlighted in Section 6; whereas the key issues of ownership, promotion and support are highlighted in Section 7. The report ends with conclusions, additional reflections and some forward-looking recommendations in Section 8.

2.0 NATO RECORDS AND IM SYSTEMS

2.1 NATO Records

NATO information is divided into classified and unclassified information. Classified information is further divided into RESTRICTED, CONFIDENTIAL, SECRET and TOP-SECRET information with varying levels of restricted access. Non-classified NATO information falls into two categories: NATO UNCLASSIFIED and Information Releasable to the Public. NATO UNCLASSIFIED information shall only be used for official purposes and only individuals, bodies or organisations that require it for NATO official purposes may have access to it. NATO UNCLASSIFIED is also subject to release procedures.[5] From a legal standpoint it is easier to handle classified than unclassified information, since the rules surrounding the former are clearer.

Operations records are identified as inactive, semi-active and active records. Inactive records come from legacy or closed operations, such as the Stabilization Force (SFOR) and the Implementation Force (IFOR). Semi-active and active records come from on-going missions such as the Kosovo Force (KFOR) and ISAF, where active records become semi-active after a certain period of time and as operational commanders see fit. Upon termination of an operation, operational records are transferred to the NATO Archives, unless they are needed for lessons learned or evaluation activities by the Strategic Command or by any other headquarters in the chain of command. Currently, inactive records pertaining to IFOR and SFOR are located at SHAPE but processed by the NATO Archives staff with the intent to transfer them to NATO Archives in Brussels shortly. KFOR information is kept in theatre, NATO Archives staff, together with KFOR Historian, have launched an appraisal process to identify information of permanent value. The intent is to transfer inactive KFOR information of permanent value to the NATO Archives shortly through the chain of command. Appraisal, retention, and disposition policies are regulated by NATO IM directives and developed by the NATO Archives Committee.

Records may be physical, but also electronic in terms of documents, reports, emails, chats, photos and videos. Moreover, and among others, the following regulations apply to NATO records:

- All NATO records, regardless of form, medium or classification level, are the property of the Alliance and are subject to the provisions of articles VI and VII of the Ottawa Agreement and/or of article XIII of the Paris Protocol;

[5] NATO document C-M(2002)60.

- Records originated and/or received by member nations while conducting a NATO mission and officially submitted to a NATO Civil or Military body are the property of the Alliance;
- Records originated and/or received by nations related to a NATO mission, but not submitted to a NATO Civil or Military body, remain the property of the originating nation and shall be managed in accordance to their relevant regulations, taking into account the requirements of the NATO Security Policy; and
- NATO shall be the owner of any document created by NATO which is based on a record from an external originator.

2.2 NATO IM Systems

NATO lacks an ability to fully identify, access, exploit operational and tactical records collectively and easily across NATO, with a view to providing immediate and joint access to objective and quantitative information for models, decision aids and planning factors. The ability to quickly search and retrieve documents across NATO military commands, non-military entities, and NATO member states through desktop computers is limited. This means that current NATO IM systems do not facilitate "seamless sharing of information and services" and other criteria (easy access; timely availability; exchange of information; effective and efficient discoverability) mentioned above.

This state of affairs inhibits operations analysis in general of legacy and current operations, and results in fragmented situational awareness and less-than-optimal analytical support to, among others, ISAF senior leaders. It also impedes knowledge sharing, meaning that knowledge is often retained personally or nationally, and not institutionally, and thus undermines NATO's institutional memory. A selection of current major NATO IM systems is detailed in Section 3.2.

Lack of immediate and easy joint access might also have indirect effects in that it is part of the reason why NATO TCNs are retaining information to ensure access. Another reason for retaining information is that NATO TCNs need to comply with national records laws and regulations. Whereas there are several reasons for the lack of information sharing, this state of affairs constitutes an additional and strong rationale for improved IM systems, and illustrates the interplay between the ability and willingness to share information.

Given that the NATO IM directives are recent and require time to interpret and implement, this state of affairs of NATO IM systems lagging behind NATO IM directives is understandable and unavoidable by definition since directives usually lead rather than lag (and formalize) current practice. The SAS-100 is a manifestation of a willingness within NATO to find tools to implement and catch up with current IM directives.

3.0 DIGITAL RECORDS REPOSITORIES: CIVILIAN AND MILITARY ILLUSTRATIONS

3.1 Civilian Digital Repositories

Digital repositories of physical documents and books are very common, sometimes very large and can readily be found within, e.g., the library sector (digitized rare/old books[6], etc.), municipalities (digitized public municipality records), the healthcare sector (digitized medical records), and genealogy databases, just to mention a few.[7]

[6] See http://www.ub.umu.se/en/search/special-collections/rara.

[7] Note that this discussion of civilian repositories is limited to scanned images of paper documents. It does not cover databases or other digitally born data.

For instance, there is the *Human Rights Documentation Initiative*[8] at the University of Texas (Austin) that encompasses the *Genocide Archive Rwanda* and the *Digital Archive of the Guatemalan National Police Historical Archive* (more than 10 million of 80 million pages scanned so far). There is also the Ellis Island Foundation that has created an online digital passenger record depository covering 22 million passengers[9], whereas the *Missouri Digital Heritage*[10] provides access to more than 6.8 million records, including the collections of the Missouri State Archives, the Missouri State Library and other institutions from across the state. As other examples, and just to mention a few, Sweden has so far digitized and made available 88 million out of 114 million population register records that cover 1895 – 1991, it has digitized all local citizen records covering 1642 – 1820 and all military staff files 1620 – 1723, among others.[11] As an additional example there is the United Nations that offers metadata and full-text search, and full digital access to UN documents, but not to operations records from peace operations.[12]

Two digital repositories of direct relevance for SAS-100 include the *Folke Bernadotte Collections*[13] (currently 2.4 million digitized pages of field records from Swedish peacekeeping units since 1956) created by the Swedish government agency the Folke Bernadotte Academy, and the *Australian War Memorial*[14] that contains digital copies of official records, private records and published collections of Australia's war experiences. Aiming at promoting research, the former example includes all available inactive and semi-active records since it was deemed more costly and time consuming to make a selection and thus make judgment calls on inclusion for each and every record, and since future information needs cannot be predicted. This means that it includes the complete recorded story of peacekeeping units. So far originally physical records and photos have been included, whereas chat, email and videos may be included at a later stage. Because of the Swedish Privacy Data Act and the occurrence of sensitive personal data in the records, only metadata search functionality has been implemented. Search results can be sorted on relevant general metadata properties, such as time, location, unit, record type, etc. This means, for instance, that timelines can be created.

Whereas civilian digital record repositories are common, they are set apart by searchability and user friendliness. Some digital record archives have weak or even no search functions, thus forcing users to navigate through folders and sub-folders instead of the searching options of metadata or full-text search; others have ineffective interfaces for displaying documents. There is thus a varying degree to which these IM systems offer *easy access*, *timely availability*, *exchange of information and effective and efficient discoverability*. Nevertheless, these examples illustrate that digital record repositories are an established, tried and tested civilian solution for search and immediate retrieval in very large records collections for in practice an unlimited community of users. The civilian approach involves the inclusion of all records (though not all records at once) instead of strategic selections, possible because the goal is one of preserving and making records available instead of addressing specific short-term and long-term knowledge needs that are difficult to identify and predict.

Costs for building digital repositories can be illustrated by the Folke Bernadotte Collections. The creation of source code for the scale-able database and web-based search interface cost around USD 100.000,

[8] See http://lib.utexas.edu/hrdi.

[9] See http://www.ellisisland.org.

[10] See http://www.sos.mo.gov/mdh/.

[11] http://www.riksarkivet.se/default.aspx?id=2102&ptid=0&refid=1020.

[12] http://research.un.org/en/docs/. Physical operations records have limited availability through the United Nations archives in New York.

[13] See http://www.folkebernadotteacademy.se/en/The-Folke-Bernadotte-Collections/.

[14] See http://www.awm.gov.au/search/collections/.

whereas digitization of individual records costs as little as USD 0.25 per page for large-scale industrial level scanning. Digitization costs will vary depending the format and quality of the records, search functionality, etc. Nevertheless, the direct costs for IT and digitization can be modest if managed in a rational manner. Digitization of records in poor condition, of odd sizes or paper qualities that do not lend themselves to automated machine scanning is much more costly and time consuming since it requires hands-on personal scanning, but this challenge is foremost to expect for old legacy peace operations. This particular challenge will thus not exist for efforts to digitize physical records from NATO operation records, many of which are already in a digital format.

3.2 NATO

The NATO Archives in Brussels provides long-term physical storage of inactive documents that are of permanent value for NATO, and some of these documents are available in digital formats. Legacy operational records can also be found in operation specific networks, and in different databases that are to different extents open and contain records with different levels of classifications. Below follows a brief presentation of major NATO and US digital records repositories of active, semi-active and legacy operations. As stressed above, the sample of non-NATO systems is illustrative rather than exhaustive, but encompasses major systems. As will be evident, these IM systems are very different from civilian digital repositories.

ISAF has created the *Afghanistan Mission Network* (AMN) that provides the common information infrastructure for the coalition members. It is a federated network consisting of a NATO provided and managed ISAF secret mission network operating as the AMN Core, with nationally provided mission network extensions of the TCN. It thus a federated network of National and NATO military networks in Afghanistan; it contains all kinds of data and formats, such as video, electronic documents, etc. ISAF is the first operation that keeps records exclusively digitally. AMN has several TCN specific search functions, but does not have a federated search service. NATO does not automatically obtain copies of records, and sharing remains possible only for certain groups.

Related to AMN is the *IJC Operational Archiving Project*. As of February of 2013, the archive contained 21.000 operational records stored without metadata as embedded documents. Of the total collection, 17.000 documents had been un-embedded and renamed to meet ISAF's naming convention. Those documents will eventually be migrated to the Document Handling System (DHS). Records are mainly collected from ISAF Joint Command HQ (COMIJC) daily briefings and include searchable metadata. Metadata tagging is carried out manually. The IJC's SharePoint portal distributes data through Sharedrive, SharePoint and DHS, and was moved to the AMN before the summer of 2013. IJC is not coordinating with the six Regional Commands (RC) concerning archiving and records keeping, and it is largely unknown if any archive by nations or US Services is being carried out at the RCs. The Headquarters Marine Corps Records Management Office surged an effort in late 2012 and early 2013 to begin to capture all US Marine Corps records. Within the scope of that effort, all servers from the latest rotation of the Marine Air Ground Task Force out of Afghanistan in March of 2013 were copied by the Records Managers. Nevertheless, there has been no coordination with the IT contractor for extracting emails from servers. Records from base closures need to be collected for processing when bases close, but resources are insufficient to carry out the necessary work. One insight is that there is a lack of guidance and expertise regarding the archiving process of IJC operational records, and that much work takes place on an ad hoc basis.

The overall goal of the NATO *Comprehensive Legal Overview Virtual Information System* (CLOVIS) is to promote easy access and sharing of curated, relevant and timely legal information within NATO. CLOVIS is browser-based tool that interconnects several databases. It is thus a search engine or search portal for other databases, but also partly a database in it own right in terms of harbouring approximately 4 GB of NATO documents based on records from NATO archives. As such, CLOVIS is a hybrid system. A current ambition is

to establish links between CLOVIS and NATO records repositories to eliminate duplicates and establish access to original documents. Full-text search will be possible when SharePoint 2010 is used as a platform, but this requires proper metadata. An important ambition with CLOVIS is to create parent-child relationships between documents so that references and connections to documents above and below are identified.

The picture that emerges from this brief overview is one of several and partly overlapping valuable IM systems, which often are mission specific, and with varying search and viewing functionality. Instead of a single NATO-wide system covering NATO and NATO TCN-owned records from all NATO operations, there are non-interconnected sub-systems with different search and viewing abilities that are tailored to specific and narrow (sometimes mission specific) needs rather than having the general goal of preserving and making records available. Hence, there is in general no *easy access, timely availability, exchange of information and effective and efficient discoverability* of relevant information. It should be stressed that nations own their part of the operational record – and will archive this themselves (in line with national policy). Hence, that there is currently not a single database for all operational records (NATO and TCN-owned) as such is thus not surprising. What is pointed here is rather the inability to search and retrieve NATO owned documents across the organization.

3.3 US Military Digital Repositories

The focus of the US *Joint Lessons Learned Information System* (JLLIS) is on joint lessons learned processes and knowledge management. It facilitates the collection, tracking, management, sharing, collaborative resolution and dissemination of lessons learned rather than the original operational records. JLLIS has a Watson data mining capability in that it searches through lessons learned documents across databases, and thereby reduces the need for metadata tagging. While not handling raw or primary mission data, JLLIS has started to collect raw survey data through external survey engines, which are currently not fully integrated with JLLIS. Instead, the output (raw data) of surveys can be added to JLLIS.

JLLIS offers broad access through different levels of classification, but the sharing of unclassified information has proven challenging. In this regard JLLIS allows owners of documents to limit access to certain users. The next step is to enable searches at different levels of classification. JLLIS is exposed to the information aggregation problem: while information in individual lessons learned reports may not be SECRET, the aggregation of information from many individually non-SECRET reports may lead to a totality of information that should be regarded as SECRET. JLLIS can meanwhile handle SECRET information. Whereas JLLIS is accessible through Internet, the security issues are handled by different layers of access that are open to users with different levels of security clearance. One lessons learned from JLLIS is that it takes time before users appreciate the benefits of a joint/shared system and mutual distrust is reduced in favour of a culture of sharing. This observation illustrates yet again the interplay between ability and willingness to share information.

The US *Complex Operations Data Development Activity* (CODDA) within the US Army Training and Doctrine Command, provides data and support for operation analysis. The objective is to establish a repository of historical operation data to support future studies and planning. The system does not yet provide direct access, Requests for Documentation (RFDs) must be submitted after which relevant documents are identified with the help of a basic metadata system and a file hierarchy, and returned to the requestors. Currently, a RFD may take anywhere from 14 to 90+ days to delivery. Note that CODDA is not a single archive/database, but a network of individual archives/databases. In contrast to – and in practice a complement to AMN – data are collected and fed into CODDA after units have redeployed and thus records have become inactive.

4.0 AN ENVISIONED NATO DIGITAL REPOSITORY: CONTENTS, FORMATS, AND SEARCHABILITY

4.1 Contents and Input Formats

It is important to move away from the current NATO operation specific approach towards a general approach that includes all NATO legacy and current operations in a single or unified approach that meets the earlier mentioned NATO IM criteria of *easy access*, *timely availability*, *exchange of information and effective and efficient discoverability*.

The envisioned NATO system would become large in terms of number of records, but civilian sector digital repositories have demonstrated that size concerns are not warranted provided that the underlying IT solution (i.e., database principle) is built with scalability in mind. Meanwhile, when systems become too large they may become slow, partly because of their size (number and size of records) and the number of simultaneous users. Allowance should thus be made for creating functionally identical sub-systems each of which contain parts of the overall amount of digital records, but meanwhile are seamlessly connected to a single user interface.

There is an unknown degree of record/information losses from inactive, semi-active and active field records in general, as caused by absence of and/or inconsistent applications of NATO IM directives. Yet, if losses are essentially overall random in terms of type of information lost, it will not matter since any operational analysis based on the remaining records/information will generate accurate conclusions. Chat information may meanwhile be especially important for tactical analysis, but it is unclear how much of the chat information that can be saved and to what extent chat information losses have been random.

Conditional on the amount resources allocated from the very beginning of establishing such an envisioned system, a choice has to be made regarding what records to prioritize. A case can be made against an initial focus on current operations, as information from legacy operations can provide valuable insights for current operations and for planning future ones. On the other hand, the current transition and redeployment of ISAF constitute a compelling reason for starting with this operation instead of legacy operations, else digital records may get lost as ISAF is scaled down.

The transition phase before an operation closes down thus poses a general challenge, as there may be too few staff to deal with records in an orderly way. ISAF serves as an example of these challenges, as there are concerns that records may be lost due to software and hardware issues and because NATO TCNs shipping out – and thus "hoard" – records they have an interest in. ISAF is also unique in that it is the first operation that keeps records exclusively digitally, and this means that it is especially easy for records to get irrevocably lost, not least since TCNs that leave the theatre may lack policies or rules for how to deal with records. On the other hand, ISAF records – NATO-owned as well as NATO TCN-owned – are easy and cheap to add to a digital repository since there are no physical records that need to be digitized. Hence, an initial focus on ISAF would make it possible that the envisioned NATO repository quickly becomes large. Other active, semi-active, and inactive records can be added at a later stage in a suitable sequence contingent on allocated resources.

When the envisioned NATO repository has been created, records from current operations can as a matter of routine be included with the shortest possible time lag. The issues concerning ISAF raised above are thus pertinent to consider during only the establishment/build-up phase of the envisioned repository.

Another important question concerns what kind of information to include: all information (just field level records, and in that case what kind of records [incident reports, data, etc.]?), and all formats (electronic [email,

chat, video, photos, voice], physical)? There are different and diverse audiences for these options, and perceived record requirements change across time. Operations Analysts (OAs) focusing on long-term strategic issues have different record needs than OAs focusing on tactical and immediate issues. Current NATO doctrine is an insufficient guide on what to include, and can moreover be expected to change in the future. The need for information on force-on-force issues is omnipresent, but there are also other knowledge needs. There is in short no consensus across OAs. Moreover, OAs in the field or just visiting should be able to pre-select and tag data worth storing, whether for immediate use or for the future.

It is difficult, costly and time consuming to not only create jointly agreed upon criteria for record inclusion, but also to make judgment calls on whether a specific record meets criteria for inclusion. Moreover, the needs of future analysts are impossible to anticipate, and decisions on exclusion may have irrevocable long-term consequences: it is time consuming to add excluded records at a later stage, and those records may also get lost. In addition, only parts of the recorded operation history would be covered. A safe and easy solution is to include all records, and in practice leave the selection of records to end-users. It appears also more rational to allocate time and resources to include all records than to allocate the same amount of time and resources to first determine what should be excluded (and run the risk that the work stalls because of disagreements on criteria for inclusion) and then examine individual documents before they are included. This all-inclusive approach is also the one applied by civilian digital records depositories that mainly focus on preservation of records rather than catering to difficult-to-identify knowledge needs.

Nevertheless, for practical reasons all types and formats of records cannot be included at the same time. It is important to start at a certain end to get the work started, and over time add operations, contents, formats and sources in a sequence that is feasible and of added value. In order to reduce the need for difficult decisions on how to sequence record inclusion, NATO can allocate more resources to enable the work to start at several fronts at the same time with regard to formats and types of information. For practical reasons it appears advisable to initiate the work by focusing on NATO-owned records, and later on add NATO TCN records when legal issues have been addressed (see Section 6) and trust as well as a culture of sharing have taken root.

Regarding record formats, civilian repositories appear to focus on including what may be labelled “documents” (whether physical or electronic) and photos, instead of emails, chat, and video. The inclusion of emails and chat communication does not appear to raise technological challenges beyond those that apply for physical or electronic reports. There are thus no compelling technological reasons against the inclusion of such input formats. Meanwhile, the inclusion of videos will dramatically – yet to an unknown magnitude – increase hardware as well as bandwidth requirements. The SAS-100 Specialist Team finds it advisable to further examine the technological and resource implications of including videos in the envisioned record depository before recommendations in this regard are made.

4.1.1 Recommendations

1) Create a single or unified approach that covers all current and legacy NATO operations:
   - Build the system with scalability in mind.

2) During the establishing phase of the IM system, prioritize the inclusion of records from NATO operations that are either being redeployed or shut down:
   - When the system has been created, records from current NATO operations should be added as a matter of routine with the shortest possible time lag.

3) Have the long-term goal of including all record format, physical as well as electronic:
   - Start with including physical and electronic documents, chat, email and photos; and
   - Examine the technological and resource implications of including videos.

4.2 Digital Output Format

Output formats for the envisioned digital record repository need to avoid becoming redundant or unreadable in the future, while allowing for searchability. The former issue is a well-known challenge for long-term storage of digital records. TIFF is currently the most secure and used format for long-term storage but because of its large size, the repository will need to batch compress/convert original TIFF files into a much smaller format, such as JPEG. Since NATO archives are using the PDF/PDF-A format, this may condition the choice of file format. Both PDF/A as well as picture formats like JPEG are popular and allow for full text searchability, in the former case directly in the document, and in the latter case indirectly through OCR software that seamlessly and in real-time interprets text in JPEG picture files. Not all records PDF/A and JPEG records – such as originally handwritten documents, photos or movies – allow for full text search, and will in that case be identified through metadata tagging/search and navigation. This point illustrates the importance of having a repository with triple search functionality, as further discussed in Section 4.3 below. In addition, since spreadsheets (Excel or corresponding) are basic tools for OAs, the inclusion of such file formats is crucial to the added value of the repository.

As mentioned above, it is advisable to further examine the technological and resource implications of including videos in the envisioned record depository before any recommendations in this regard are made. In this regard it is important to carefully identify a suitable movie output format.

4.2.1 Recommendations

4) Select PDF/A and/or JPEG as output formats for documents, chat, emails and photos; and

5) Assess suitable output formats for movies before decisions on inclusion are made.

4.3 Searchability

The envisioned repository would be well served by the three main search interfaces – navigation, metadata keyword search, full-text search – each of them having strengths and weaknesses. Full-text search functionality should also be combined with a Watson-type searchability.[15] Records will thereby be locatable by navigating from folder to sub-folders, to sub-sub-folders, and so on; through metadata search; and through full-text search. Different users may from time to time prefer different approaches, and this triple functionality provides the communities with a menu of choice. Users who know where a specific record is located will find navigation to be the fastest approach; users who lack such knowledge, or are searching for non-specific documents, will prefer meta-data search or full text search: hand-written documents, photos and movies that are not locatable through full text search can only be located through navigation or metadata search. In addition, full-text search and metadata search should be possible to combine in a single search event, as it enables users to further narrow down search results. Finally a record repository should have a functionality where datasets (building on the original records sources) or similar by-products from the repository can be uploaded and made searchable for the wider community of users.

[15] The US JLLIS experience shows that an advanced Watson search engine can be procured for approximately USD 200.000.

Search results should be sortable on relevant general metadata properties, such as time, location, unit, record type, etc., or full-text search terms. Thereby, for instance, timelines of identified records can be created. In addition, search results should be sortable on the widely applied "popularity" property as applied by Google and other search engines, i.e., how often records have been retrieved by users.

Metadata search functionality as well as navigation require indexing, which means that searchable tags are added in terms of time, location (different geographical/administrative units, or even geographical coordinates/grid information), unit, HQ, sub-unit, battalion, company, platoon, section, document type, serial headline, etc. This creates also the very initial record structure that allows for not only metadata search but also navigation as well as for ordering the physical/original records. Large-scale indexing can be carried out fast through batch indexing, which means that all records in a single "volume" (of any size) are automatically indexed in an identical manner.[16] Indexing software functionality should be inbuilt in the repository database to allow for seamless indexing of records, and should be tailor-made with predefined entrance fields for the predetermined metadata tags categories (e.g., "unit", "date", "battalion" and "serial headline") to promote comparability among records and consistency of tagging.

Different individuals use different methods and concepts to label records and files: sometimes the same type of records has different labels; sometimes different types of records types have identical labels. There is no common standard consistently applied across archived and non-archived documents/information, and this is a general issue that applies not only to NATO but also to other organizations and countries. Attempts to reclassify NATO operational records according to some typology are likely to be challenging, as any new batch of records may create a need to revise an existing typology. A practical solution is to be aware of these metadata inconsistencies, to keep the current classifications when adding metadata, and for users to employ several search terms to increase the likelihood that the desired record is covered. Preserving the original classification means also that a digital repository will mirror the physical/underlying archive, which is valuable if a user desires to locate a digital record on the basis of information in the physical/underlying archive (for instance through navigation in the digital repository) or locate physical records on the basis of the digital repository.

One way to ensure that metadata tags are applied consistently throughout an operation is to integrate their application the start of an operation. This requires something like a permanent branch for (data) analysis from the start of an operation. An alternative solution avoids burdening operations with this task and instead gives the final metadata tagging work to the administrator of the envisioned record repository. Hence, the final ordering of records and general labelling is carried out by entities that receive field records, and it increases consistency of metadata tagging. Nevertheless, for NATO purposes the practicality and feasibility of field-based metadata tagging versus metadata tagging by an envisioned repository owner needs to be further examined.

4.3.1 Recommendations

6) Implement navigation, metadata keyword search, and full-text search functionality:
   - Metadata search and full-text search shall be possible to combine in a single search event.

7) Datasets (building on the original records sources) should be possible to upload and made searchable for the wider community of users;

8) Search results should be sortable on general metadata properties, such as time, location, unit, record type, etc., or full-text search terms. In addition, search results should be sortable on "popularity";

[16] A case in mind is the *Folke Bernadotte Collections*, in which a single person indexed 700.000 pages in less than 10 weeks in 2013.

9) Use original records classifications when adding metadata to digital records; and

10) Examine the practicality and feasibility of field-based metadata tagging versus metadata tagging by an envisioned repository owner.

5.0 DOCUMENT CLASSIFICATIONS AND SECURITY CONSIDERATIONS

A web-browser-based user interface for access should be considered as it is commonly used in military and civilian digital archives and IM systems, even those that are classified as SECRET. It is an attractive solution as it avoids investments in – and support costs for – client software, and allows access through standard PC and Apple computers.

Handling and sharing of classified data need to comply with NATO security policies. One option in that regard is a system comprising separate layers of data access (e.g., separate layers for users with different access rights; separate layers for different record classifications). While layered access for different record classifications and users with different levels of security clearance may be considered and has been successfully implemented in JLLIS, it appears relevant to consider whether users who lack security clearance up to the level of SECRET are really the target group of the repository. It may also make the system more complex and expensive to create, would entail costs for properly metadata tagging records with security classifications, and run the risk of enabling access to SECRET information – as caused by the non-SECRET information aggregation problem – to individuals with lower security clearance.

An option is thus to classify the repository as SECRET to enable access to non-classified records, and classified records beneath the level of COSMIC TOP SECRET. It would address the information aggregation problem: because of the envisioned size of the repository, the aggregation problem can be expected to be a practical rather than theoretical concern. It would also target primary user groups and reduce complexity and costs for creating as well as running the system. Access should by default be restricted to NATO countries only; access for non-NATO countries may be managed through bilateral MoUs. In this regard an issue to consider is whether active and semi-active records from current operations should be remain accessibly for TCNs only in terms of, e.g., “releasable to ISAF.” Implementing similar release rules for active and semi-active records would serve to undermine the goal of NATO-wide sharing of the most topical records and reduce the added value of the envisioned repository.

5.1 Recommendations

11) Employ a web-browser-based user interface;

12) Classify the repository as SECRET; and

13) Restrict access by default to NATO countries, while access for non-NATO countries may be managed through bilateral MoUs.

6.0 LEGAL CONSIDERATIONS

An important issue is the non-harmonized laws of NATO nations on privacy, copyright and national security. NATO records are meanwhile not protected by privacy laws. There are thus national laws against sharing. A further complication is that sometimes it is unclear whether NATO or a certain NATO TCN is the originator

and owner of a record although it originated at a HQ. This issue sometimes arises when NATO TCNs belong to, e.g., Operation Enduring Freedom (OEF) or ISAF, but at the same time have their own operations. In these cases the issue sometimes becomes whether the record was created as part of the NATO chain of command, or the national chain of command. Current NATO IM directives make it difficult to determine ownership in such cases.

One possible solution involves that all NATO records are by default shared by all NATO members, while sharing of NATO TCN records is regulated by multi-lateral or bi-lateral MoUs. Such MoUs could be made compulsory among all NATO TCNs and thus a standard feature of all NATO operations, and would involve consent to share national operations records, provided that national security laws and privacy laws are upheld. Some level of national record check will thus be necessary before sharing can take place, but the MoU would still mean that the default position will be one of sharing, it would foster and enable sharing, while at the same time retain national control over records. It is also important to establish an agreement from all NATO TCNs on minimum requirements for sharing and to ensure that the process is in line with the new NATO policy on ensuring access and exploitation of digital formats.

Related to this is that the envisioned repository should include a user interface that allows individual NATO TCNs to control and grant NATO/operation wide access of their records. Thereby, the envisioned system can be created before a strong and widely applied culture of sharing has arisen, and countries can have control over – but also an easy way of approving access to – their records.

In addition, it is important to revise current NATO records policies to reduce lack of clarity regarding ownership, and to improve the consistent application of those directives. The latter is also important for security reasons, since handling and sharing of classified records need to comply with NATO security policies. However, NATO's security regulations concerning handling and sharing of records will always require a degree of interpretation before they are put into practice. A digital records repository would therefore be well served by the establishment of a standing legal adviser and a legal council.

6.1 Recommendations

14) Restrict access to NATO records by default to NATO countries; access to NATO TCN records should be managed through bilateral MoUs between the particular NATO TCN and NATO;

15) Include a user interface that allows individual countries to control and grant document NATO/operation wide access of NATO TCN owned records;

16) Consider the feasibility of revising NATO records policies to reduce lack of clarity on record ownership; and

17) Establish a standing legal adviser and a legal council.

7.0 OWNERSHIP, SUPPORT AND PROMOTION

For practical reasons, a single or unified NATO repository is preferable over federated separate/multiple repositories (i.e., a federated network of NATO repositories). A single administrator/custodian ensures coherency, compatibility, sustainability, consistency, clear ownership and a long-term strategy and stability in terms of contents, functionality, promotion, securing viability of digital formats and access. Moreover, a unified system with a single access point and user interface avoids that users need to search for information across networks and systems that have unique properties in terms of metadata tagging, searchability, access and interface. It would also allow for central registration of access as well as coherent and timely re-classification (downgrading and declassification) of records. A single system will also by definition eliminate the risk that

multiple yet intentionally complementary systems end up overlapping and competing with – instead of complementing – each other.

Active records may be managed in repositories at field level strategic commands, but since it is difficult to draw a line between active and semi-active records, and because of the aforementioned advantages of a single administrator/custodian, it is not an ideal solution. SHAPE, NCI or ACT appear instead to be the best options for managing a repository for digitized semi-active and active records, whereas NATO Archives is best suited for managing a repository of digitized inactive records which are already under its administration.

These entities have clear responsibilities and know-how, but currently need increased resources and would need to work together with information specialists in the command structure to manage the envisioned repository. While there are thus two general administrators/custodians, there needs to be a single access point in order to facilitate ease of use. To coordinate the work of the two complementary administrators/custodians, there is a need to create a joint steering group with a joint vision and goals.

Ownership (administrators/custodians) and location of the repository must not necessarily be identical, but will for practical reasons be the most beneficial solution as it ensures easy and close day-to-day and face-to-face informal communication and cooperation among concerned individuals.

7.1 Recommendations

18) Create a single common repository with two administrators/custodians, involving SHAPE, NCI or ACT for semi-active and active records, and NATO Archives for inactive records; and

19) Co-locate repository administrators/custodians and the repository.

8.0 CONCLUSIONS AND FINAL RECOMMENDATION

The envisioned NATO digital repository that meets the objectives of SAS-100 would fulfil the visions of current NATO IM directives. Civilian digital repositories have demonstrated that concerns in terms of the number of files are invalid, and that search, retrieval and viewing functionality can be fast and user friendly. The envisioned digital repository would also in practice share many features with widely used civilian repositories. The challenges facing the envisioned NATO digital repository are neither foremost technical nor financial, but concern interconnected legal issues and trust. Increased trust may be expected to be an outcome – rather than source of – the envisioned repository, whereas an MoU approach as described above may to a sufficient degree address legal issues but also increasingly and stepwise improve trust. The envisioned repository may also promote the development of a culture of sharing, may lead to a change in and a more consistent application of NATO IM policies and directives, and generate improved and more coherent archival skills and application at the theatre level.

The envisioned repository will not meet all demands and preferences regarding contents from the beginning, but can be made more comprehensive over time. The work has to start in some corner. Whereas not all records and format will become immediately or even ever accessible through the envisioned repository, it would still constitute a huge leap forward for NATO as compared to doing nothing. It is important to stress that the issues facing NATO in terms of lost and incomplete records are also faced by, among others, the UN. Such issues should thus not be overstated, and are not decisive for whether the envisioned repository should be established.

One issue that was not covered by the SAS-100 objectives, but nevertheless is important, concerns access to non-NATO TCN records by NATO, as well as access to NATO records by non-NATO TCNs. Non-NATO

troop contributions have constituted important elements of NATO operations, and the countries in question have also created records of relevance for operational analysis of past and current operations. It appears valuable for NATO to consider creating an additional SAS Specialist Team to assess:

a) The added value of access to non-NATO TCNs records, and by non-NATO TCNs to the envisioned repository; and

b) The legal and practical issues involved in such joint access.

A follow-on SAS Panel may therefore be warranted to address this any other issues that this report may generate questions on. A resolution of the issue of inclusion of non-NATO TCN records is meanwhile not decisive for a decision whether to create the envisioned repository. The issue can be resolved at a later moment in time.

8.1 Recommendation

20) Create a SAS Specialist Team to assess the added value of access to non-NATO TCNs records, and by non-NATO TCNs to the envisioned repository, and the legal and practical issues involved in such joint access.