A Semantic Approach to Information Management and Decision Support: An Application to Humanitarian Demining Operations

Linus Gisslen and Anders Törne
Swedish Defence Research Agency (FOI)
Stockholm, Sweden & Linköping, Sweden
{linus.gisslen, anders.torne}@foi.se

***Abstract*—This paper discusses an approach to information management and evidence based reasoning in the realm of hazardous contamination, particularly humanitarian demining. It proposes a system called HAIMS (Hazard Information Management System) to support the user when classifying and ranking the priority for further information gathering and/or operations (operations in this context can be mine clearance, technical surveys or non-technical surveys). This includes the use of semantic information system, evidence assessment and community impact from hazard contamination.**

Keywords—Information management system, knowledge management, decision support, semantic techniques, demining operations, evidence assessment, humanitarian demining, social impact factors, situation assessment.

I. Introduction

The effect of land mines cannot easily be overstated in terms of human suffering. Mine accidents result in immeasurable personal suffering and demand extensive society resources which could otherwise be used in a more productive way. Furthermore, land mines do not only cause direct suffering in terms of human lives and injuries, but they also impact on the long term development of a society, productivity and education level [1]. Landmine contamination can affect the society in many ways: the presence of a mine threat can for example block children's physical access to school and adult's access to employment. The overall society risks to be denied access to medical facilities, water access, food sources, etc.

To mitigate the effect of land mines much effort has to be spent to locate and possibly remove the hazards. However, not only do identified physical hazards play a role in the development of a society but also a (possibly false) notion that hazards are present in a given area could negatively impact the standard of living and productivity. To encourage the local population to use land for food or other production, they need to be convinced that no hazard exist in the area. As important as it is to locate possible hazards, it is equally important to prove the absence of hazards. For both processes it is crucial to gather evidence of hazards, prioritize demining efforts and localize hazardous areas and hazard-free areas to the best quality possible. Therefore information technology advancement has been identified as one of the most important improvements for the demining community [2]. As a demining clearance operation is very costly, even simple marking of the identified contaminated or suspicious areas can be a cost effective way to increase the productivity of society [3].

Humanitarian demining programs and regional operations are established based on the consideration of the impact of the hazards of war and terrorism on an area and its people. Once the need for remedial action is evident, information is gathered in a survey of the area and the suspected hazards to define the extent and, if possible, the nature of the hazards. The acquired information will cover relevant administrative, cultural, demographic and logistical factors and other information necessary to plan and deploy a demining program or operation. The process of assembling information remotely and at a safe distance plus the assessment and recording of that information is known as a Non-Technical Survey (NTS). This process will continue as further data is collected. The NTS is both anticipatory and responsive. NTS can be at varying levels: it can be country-wide, regional or local in coverage depending on the stage of the country's demining process.

The decision to release land from suspicion of mines has historically been an assessment partially based on personal experience of the demining personnel. This assessment is often conservative and results in land being overly classified as "suspicion of mine presence" rather than the release of the land. This is perhaps as expected when a released contaminated field could lead to personal disaster. However, a conservative judgement also leads to a large amount of areas being wrongly classified as suspected hazard areas (SHA). This leads to either large efforts in on-site manual searches in uncontaminated areas, or land being left that cannot be used for production or passage by the local population.

This paper presents a concept to improve this situation by providing decision makers with support to assess the current situation with respect to hazardous threats (the importance stressed in paper [4]). The concept addresses the information management part of humanitarian demining and subsequently the planning of and decision making in these kind of operations.

II. Related Work

Information management is the effort to collect and manage information between sources and users. The paper [5] describes how evidence can be managed by using semantic annotations via semantic MediaWiki. The demining operation is truly a task

978-1-4799-8657-6/15 $31.00 © 2015 IEEE
DOI 10.1109/EISIC.2015.34

of evidence assessment and decision making on very limited and/or weak information. Theoretical work in this direction can be found in papers [6, 7]. Within the context of demining the most widely used system is the information management system for mine actions (IMSMA) [8]. IMSMA is an UN-approved system supporting humanitarian demining. IMSMA is designed to store data from mine actions. It supports the user in reporting, monitoring and decision support. It allows for advanced information management using dynamic forms and customized relations between objects. It also contains support for calculating community impact scores (see *II.A* for definition of community impact). This scoring tool allows the user to perform impact analyses for hazard areas using the data stored in IMSMA. The factors included in the calculation needs to be selected and weighted by the user. The total impact is a function of factor weight multiplied with a point value of the factor. The score is then the sum of all factor products. Other software demining applications like MASCOT [9] also support impact calculations.

IMSMA does not have support to calculate the probability of hazard contamination. To help with calculating the probability of hazard contamination previous work has developed a method to quantify this number. In the paper from Larsson et. al. an algorithm is described that assesses the importance of each individual piece of evidence [10]. Each evidence (E_j) is weighted with the credibility ($c_j \in [0,10]$) of each source of information and the importance ($w_{ij} \in [0,10]$) relative to other information. Evidence could either support the presence of mines ($M_j=1$), or support the absence of mines ($M_j = -1$). The equation that calculates the value for area S can thus be written as:

$$V(S) = \sum_{j=1}^{J} M_j w_{ij} c_j$$

The value is not a probability but a quantitative number: if the value is below a certain value the recommendation is Land release (i.e. the land is considered free of hazards), if it is over a certain threshold a clearance operation is recommended, etc. The corresponding thresholds can be found in the article [10]. This approach does not take into consideration the cumulative effect of having several evidence. Two evidence together are a stronger indicator than the two of them treated separately and multiplied. Amongst other things a solution to this will be presented in this paper.

III. PROPOSED METHOD

In this section we propose a model for managing information related to hazards called HAIMS (Hazard Information Management System). The layout of this section follows: first a set of specific words will be defined (*A*) not to confuse the reader as in other realms these words might have a different meaning. After the main objectives (*B*) are defined, then the concepts used (C) followed by discussion on Information quality (D) and Decision support (E). (F) summarizes the section.

A. Definitions

- Indicator: An observable factor that supports either a hypothesis or its negation. An indicator can be either an event, a physical or immaterial object including its state or an activity.
- Evidence: A specific observation by one observer of an indicator is called evidence.
- Direct evidence: In the demining context it is often a direct observation of hazards, or an observed direct effect like an observed explosion by a credible observer. It can be interpreted as synonym to proof.
- Hypothesis: A belief held without proof. In the context of demining a hypothesis is typically about hazard presence.
- Information Object (IO): An IO is any describable object that can have attributes. Examples can be infrastructure, person, mine accident, event, hazard area (see below), etc. Each IO has a specific set of associated attributes depending on the object type. See section *Information object* for more information.
- (IO) Attributes: describes an IO. The attributes depend on the IO type. Examples are date, location, coordinates, usefulness, population, age, etc.
- Hazardous contamination: any kind of contamination of dangerous materials like explosives, booby-traps, chemicals, toxics etc. In this paper we focus on the hazards like land mines that are explosive remnants of war (ERW).
- Hazard area (HA): An area that potentially could contain contamination in the form of hazardous objects like land mines, bombs, other unexploded ordnance (UXO), booby-traps, etc. See section *Information object* for more information.
- (Community) Impact: The negative effect that a hazard area has on its surrounding communities. It is commonly referred to as impact factor in the demining community. In this paper it is separated into two parts: a social impact (which is an indicator how a HA effects the population) and a functional impact (which is an indicator how a HA effects the productivity/usefulness of a land).

B. Main objectives

The following important aspects derived from the problems described in the *Introduction* are addressed in this paper:

1. How to correctly assess social and functional impact on a society affected by hazardous contamination.
2. How to support evidence gathering and evidence assessment, and calculate the probability of hazard presence.
3. How to design the decision support to maximize the end-user gain.
4. How to ensure overall information quality. The information needs not only to be filtered when it is recorded into the system, but also when it is retrieved from the database and how it is displayed.
5. How to present an operational picture which needs to be clear and intuitive.
6. How to reduce the complexity in the information management so it is accessible to non-experts. I.e. how to construct a system which is data rich and complex but is perceived as simple to use in the eyes of the end user?

C. Concept

To support informed decisions in the domain of hazard contamination a concept is presented to help with the management of information. It has three main components:

- A database: This is used to store all the information gathered and allow the user to retrieve the desired data when it is needed. To support informed decisions with focused information the database contains semantic support.
- An interface: The best way for the user to access the information in the database is via an intuitive interface. The interface is also used to enter information into the database.
- A decision support system: this is the core of the concept. It serves as a filter and aggregator of information between the database and the interface. It ensures that the information sent from the database to the interface is relevant and refined. Refined means added value, for example a calculation from the stored information objects in the database is to refine the information. It also supports the user in storing the data to maximize the usefulness of the added information.

To illustrate the structure of information we can view information in three refinement levels:

1. The untreated data mainly added to the database by a user or by automated imports. This is usually in the form of an information object and its attributes.
2. The information obtained from the semantic implementation by using the first level information. E.g. distances between locations and other relationships between objects. Also aggregated information like population within a region belongs to this category.
3. The information obtained from making calculations on and/or refining the information in the other two levels. It could for example manifested as impact factors and probabilities.

As many software support within the demining community rely mainly on the first level of information, except for special cases like impact scoring tools. The concept HAIMS presented in this paper focuses mainly on the two higher levels of information refinement.

1) Semantic functionalities

A need for a way of refining information has been identified without explicitly requiring an input from the user. The semantic techniques [11] allow the use of semantic annotations which give the functionality of a collaborative database. This allows for the automatic retrieval of information not explicitly recorded. Examples of queries (in demining context) could be:

1. List all minefields within a radius of 2 km that have or have had mines at a depth >10 cm, and has soil type Sand.
2. List all mine accidents that are less than one month old in region X.
3. List all hospitals within a radius of 10 km from point X and list their distance and function.
4. List all evidence within 500 m that indicates that there is hazardous contamination there; the coordinate error must be less than 10m.
5. Retrieve how many people live within 10 km from coordinate X.

The queries can answer questions which otherwise would not be feasible to answer without semantic annotations. The semantic techniques allows for gathering information automatically without requiring the user to do manual work as the queries can be hidden away in presentation templates. The semantic technique allows for an attribute of an IO to be interpreted by the system as properties. E.g. the coordinate attribute can be encoded and give the relative distance between different objects. An example - an often desired piece of information to have when planning operations for a hazard area (HA) is the nearest hospital. In traditional systems the user has to manually fill in the hospital for each HA (which is often the same for several HA) and the distance then has to be calculated or retrieved in some other way. With the semantic approach it is only required to create one instance of the hospital and then the application will automatically retrieve the closest hospital and its distance for each HA (compare query 3). Similarly, the HA can have relevant evidence connected to it by automatically gathering information in the region, e.g. from other nearby hazard areas or previous mine accidents, conflicts, etc. (compare query 4). Information how many people the HA is affecting is easily retrieved with one query (compare query 5). Semantic functionalities will be used later when the calculation of social impact, contamination probabilities and other decision support features are discussed.

2) Information objects

An information object (IO) is a container that holds attributes about an object and is mainly used to store information. Below we describe information objects (IOs) that are of particular importance in the presented concept.

Hazard area (HA): This is the main IO to store information about a (potentially) hazardous area. It contains attributes like soil type, area polygon coordinates, metal level, etc. - things that describes the features of the area itself. It will help in deciding which tools to use, work required (for demining), and so on. There are also other more abstract calculated attributes like *probability of hazards*, *social impact* and *functional impact*. The probability of hazards is calculated from evidence and the social and functional impacts are gathered from information about infrastructures.

Evidence: The evidence IO contains attributes about an *evidence*, e.g. evidence type, evidence reliability, age, HA connection, etc. The concept of evidence is described in more detail in section *Evidence and indicators*.

Infrastructure: The infrastructure IO can have several uses. Firstly, if the infrastructure has a social function it has a usability quantifier (*u*) indicating how useful it is to the society. If the infrastructure is proximate to a hazard area (i.e. affecting access or usage) this infrastructure will contribute to the negative impact the HA is having. The infrastructure information can also be of use in the planning the demining operation, for example infrastructures like Safe access routes, Neutralization sites, Hospitals, Police stations, etc.

Settlements: This IO is similar to infrastructure but has a population (p) like cities, towns, villages, refugee camps, etc. Similarly to infrastructure the value of p will decide how much a HA is negatively affecting the surrounding region.

Information about other domain specific information objects can also be stored, e.g. demining tools, personnel, Personal Protective Equipment (PPE), etc. They all have attributes that makes them searchable. For example, a Demining tool might be suitable for certain soil types, metal levels, etc. The semantic calculation allows for easy search in the database for finding well suited tools when deploying a demining operation like a technical survey or a clearance operation.

D. *Information quality*

The information and data quality is naturally one of the big challenges in any information management (IM) system, and especially in a safety critical domain realm like demining where lives are at risk. Any IM system is only as good as its data and if not each piece of information about an IO is trustable the whole IO can be useless. For example, a very well recorded mine accident can be unusable if the coordinates recorded were measured with an unknown (or potentially error prone) method: who would risk their life on a coordinate which could be erroneous? Therefore, in order to make information useable within a system that has several users and where data is shared, it is crucial to motivate each piece of data - especially for safety critical information. For this purpose we propose a directed input form to record the data where each piece of critical information has to be validated. We identify and describe two critical information types: The coordinate attribute (*1*) and the Evidence IO (*2*).

1) Coordinates

The coordinate attribute set is designed to increase user confidence to make a decision regarding the coordinates of the IO. The decision can either be to accept the coordinates, or to reject them if they do not meet the expected quality. A summary of the coordinate attribute set is listed here:

- The coordinate position number(s) (e.g. N33, E33). Could be a single coordinate (point) or a set of coordinates (a polygon describing an area).
- The coordinate system (e.g. WGS84).
- The *reliability* of the source: A qualitative measure how trustworthy the source is (e.g. Reliable, Not reliable).
- The *accuracy* of the source: A quantitative measure how accurate the source is. Can be described as a quantitative (e.g. *less than 10 m error*) or qualitative (e.g. *Approximate*).
- How the coordinates were obtained. A description of the method used to obtain the coordinates (e.g. differential GPS with reference point).

The idea is that coordinate attribute set is identical for all IOs in the application to help with the standardization of entering data about IOs and therefore ensure data quality.

2) Evidence

In the context of demining a hypothesis can for example be “Area X contains hazard contamination”. Evidence are used to decide if a hypothesis is true or not. The hypothesis is then used to classify an area, e.g., hazard free or hazard contaminated. As previous mentioned evidence could either be in support a hypothesis, or be against a hypothesis. E.g. *Crater* would support the above hypothesis in this case. *Cultivated land* would not support the same hypothesis. Naturally several evidence can contribute to the total likelihood of a hypothesis being correct.

In the demining community one type of evidence is *direct evidence*. E.g. observation of hazards can be considered direct evidence of hazard *presence*. On the other hand an area cannot be released from suspicion of mines based only on no observations have been made (mines might be hidden). A proper mine search (a technical survey) according to national criteria where no mines have been found is considered direct evidence of *absence* of hazards. A direct evidence alone is sufficient to classify a land.

The other type is *indirect evidence*. It can support a hypothesis but alone it is not enough to classify a hazard area. However, several evidence pointing in the same direction could strongly support a hypothesis. Examples of indirect evidence can be “productive land in use”, “craters”, “satellite images indicating man-made objects”, “fighting in area”, etc. Indirect evidence together with impact analysis may be used to prioritize operations in a situation where many areas are suspected to contain contaminations.

Important to note here is that evidence should be analyzed in relation to a position. A specific *evidence* can simultaneously both be ‘Direct’ and ‘Indirect’ depending on the location relative to a given area. Consider *Mine observation* as an example: in the area where it occurred it can be considered a *direct evidence* for that particular area. On the other hand, for the neighboring areas it can be indirect evidence for hazard contamination as mines usually do not come alone. Adjacent evidence increases the probability of the hazard hypothesis, therefore it is important to be able to search in the database for nearby accidents or confirmed hazards areas etc.

A method to record evidence is proposed so that it meets the (extremely high) quality requirements of the demining community. The evidence assessment is divided into two main factors. First, the assessment of the information itself. It is designed to answer questions like: Is the information true? Is the description correct? Etc. Second, the assessment of the interpretation answers questions like: Does it prove hazard presence? If so, what hazard? If not, how strongly does it support/not support the hypothesis?

To assess the information part, a list of attributes needs to be recorded in order to make up a complete and meaningful evidence IOs. The most important attributes are listed here (Omitting the coordinate, see previous section):

- **Date**. The age of the evidence is crucial because it puts it into a sequential/time context.
- **Type of statement**. What type or fact is it? E.g. ‘Crater shown in an aerial photo’ or ‘On-site observation of mines’.

- **Information credibility**. This assesses how likely the statement is to be true: If the evidence is a mine observation then a mine expert will have a higher credibility than a layman. Furthermore, a deemed unlikely evidence might have lower credibility.
- **Information reliability**. Similar to *Information credibility* but rather answers questions like: How biased is the testimony or report of the information source? Does the person have personal gain to say one thing? Is the evidence very old so the testimony is outdated or misremembered?
- **Source**. Where does the information come from? E.g., interview, observation from the reporter, archive, tool measurement.
- **Information conclusion**. Is the statement accepted as reliable by authorized personnel, so it can be used in further analysis?

When, and if, the statement is accepted, the user needs to define what the fact implies. For example, what does the fact that there are craters in an area imply? A discussion about what a certain fact indicates in the context of demining has been analysed by Bajic [12]. The most important implication attributes for evidence IO are:

- **Evidence direction**. Does it support or disprove a hypothesis of hazard contamination?
- **Evidence type**. Direct or indirect evidence?
- **Evidence strength**. If the evidence is an indirect evidence the strength is given by two values: the first expresses the frequency to find this evidence in a field *with* hazard contamination. The second expresses the frequency to find this evidence in a field with *no* hazard contamination (i.e. false positives). These numbers are later used to calculate the probability of contamination.
- **Hazard characterization**. (If applicable) Type of hazard? Number of hazards? Maximum depth? Etc.

After these sets of questions have been answered a first evidence assessment should be possible. The evidence will also include the coordinate attribute as the location of the hazard is crucial.

E. Decision support

As stated before semantic techniques give a good platform for decision support. Here is presented how the derived information from semantic queries can be used to improve the decision support and information refinement.

1) Impact calculation

Almost every suspected or confirmed hazard contaminated area has an impact on the surrounding. To quantify this affect each hazard area can be assigned an impact score. This number can be interpreted as the positive impact a removal of the contamination would have. Therefore every infrastructure and geographic object around an area can potentially contribute to the beneficial impact of clearing and releasing that area for normal use. An impact algorithm needs to satisfy a number of requirements:

- It should be intuitive and reflect the view of experienced users.
- It needs to be simple and easy to follow (i.e. no "black-box").
- The impact factors need to be dynamic so the user can try different initial conditions and do their own analysis.

In this model the impact is calculated with two aspects in mind: the social impact and the functional impact. The social impact is a function of the number of people affected and their inverted distance to the area. The higher the number of people, and the closer they are, the higher the impact. The social impact includes populated areas (i.e. settlement IOs) such as villages, cities, refugee camps etc. The required data is population (p) and coordinates; both are given in the IO attributes. The contribution to the impact I for each hazard area can be written as:

$$I^{Social}(d,p) = \frac{p}{d}C \qquad (1)$$

Where p is the population of the object, d (with a lowest value of 0.1 to avoid division close to 0) is the calculated distance in km between the settlement and HA and C is a normalization constant. The total value of the social impact is the sum of all the social impact objects:

$$I_{Total}^{Social} = \sum_{n=1}^{N} I_n^{Social} \qquad (2)$$

Where N is the number of settlement objects within a given maximum distance.

The functional impact involves *infrastructure* IOs such as crop fields, roads, schools, hospitals, water resources, and other infrastructures with social functionality. Similarly to the social impact, the functional impact depends on the infrastructure importance (u) divided with the distance to the hazard area. As earlier described the required attributes for an infrastructure IO are its estimated usability (u) to the community and coordinates. The functional impact factor can be written as:

$$I^{Func.}(d,u) = \frac{u}{d} \qquad (3)$$

Similarly, the total value of the functional impact is the sum of all the functional impact objects.

$$I_{Total}^{Func.} = \sum_{m=1}^{M} I_m^{Func.} \qquad (4)$$

Where M is the number of infrastructure objects within a given maximum distance.

The two factors can either be treated separately or together depending on usage. The total impact for a hazard area can then be written as:

$$I^{Total} = I_{Total}^{Social} + I_{Total}^{Func.} \qquad (5)$$

The normalization constant C in Eq. (1) lets the user set the relative importance between the social and functional impact. The value of I^{Total} should be seen as a relative attribute to help in prioritizing between areas for operations like clearance, surveys, mine education, etc. The equation is designed to use stored IOs in the database: the system will gather all the information needed from the database and do the calculation without interaction of the user. It supports unbiased decisions by the user, and simplifies the ranking process when prioritizing important areas, especially when they are numerous.

2) Probability calculation

To assess evidence and estimate probable hazard contamination is difficult for even the most experienced deminer. There is a lot of information to take into consideration and every situation is unique. As previously pointed out there is a risk of conservative bias in the demining environment because the resulting effects of erroneous decisions are devastating. Furthermore there is decision asymmetry – it may be perceived easier to take a decision to keep a certain area suspicious than it is to release it. This might result in unnecessary large areas being classified as suspicious. Other factors like overestimating/ underestimating the importance of certain data, a tendency to search for evidence that confirms a biased belief, prejudice and group thinking can also negatively affect the conclusion.

We propose a method to help estimate the probability of hazard contamination to support the decision in the classification of lands. The algorithm sums up all the evidence, both for and against hazards, and calculates a probability value. The algorithm calculates values based only on existing evidence IOs in the system. To calculate the relative probability of hazards we use Bayesian theory/rule with multiple variables [13]. The Bayes' Rule for multiple variables can be written as:

$$P(A|X_1, X_2, \dots, X_n) = \frac{P(A)P(X_1, X_2, \dots, X_n|A)}{P(X_1,X_2,\dots,X_n)} = \frac{P(A)\prod_{i=1}^{n} P(X_i|A)}{P(A)\prod_{i=1}^{n} P(X_i|A) + P(\neg A)\prod_{i=1}^{n} P(X_i|\neg A)} \quad (6)$$

In this context the rule can be used to calculate the probability of hazards in a given area: assume *A* is the hypothesis (e.g. *Area Y has mines*) and X_i is a variable/evidence (observation of explosion craters). From this follows:

- *P(A)* is a quantified answer to the question: *What is the probability of hypothesis A to be right.* If the hypothesis is *A*={Y has mines} the *P(A)* is the answer to: *What is the probability of mines in the area Y?* The value is expressed in percent. For example if a region has 2% mine field areas, the value of *P(A)* is 2/100.
- $P(A|X_i)$ is the probability of hypothesis A to be right given X_j. I.e.: What is the probability mines if there are explosion craters in the area Y?
- $P(\neg A) = 1 - P(A)$ is the probability hypothesis A being wrong: What is the probability of no mines in area Y?
- $P(X_i|A)$ is the probability of X_i if hypothesis A is right. I.e.: What is the probability of explosion craters if there are mines in the area Y?
- $P(X_i|\neg A)$ is the probability X_i if hypothesis A is wrong. I.e.: What is the probability of explosion craters given there are no mines in the area Y?

With the formula it is straightforward to add further evidence and to calculate the mathematical probability. For example, to express the probability for an area that has both craters and dead animals (X_2) we can write *P=(A|X₁,X₂)=P(Has mines|Has craters, Has carcasses)*

$$P(A|X_1, X_2) = \frac{P(A)P(X_1,X_2|A)}{P(X_1,X_2)} = \frac{P(A)P(X_1|A)P(X_2|A)}{P(A)P(X_1|A)P(X_2|A)+P(\neg A)P(X_1|\neg A)P(X_2|\neg A)} \quad (7)$$

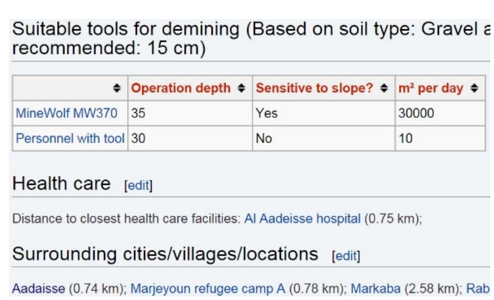
The Bayes' equation allows having multiple evidence that accumulates their effect, or "works against" each other. If X_2 is evidence against the hypothesis, the P will be lowered accordingly. Unlike previous approaches, the calculated value is the mathematical accurate probability for hazards assuming the input values are correct.

The limitations/drawbacks in using this kind of Bayesian approach are:

- There is no time aspect when calculating the probabilities. Neither chronology nor aging factor is considered.
- The evidence requires that a user is estimating the $P(X_i|A)$ and $P(X_i|\neg A)$ (expressed in frequency of being correct and/or incorrect). This can be solved by setting standard values for each type of evidence. However, it might be region dependent and therefore hard to approximate.
- The evidence inputs need to be independent from each other. If several observations of exactly the same indicator (the same carcass, for example) is entered that the evidence is not independent. This effect can in the current model only be mitigated by inspection and acceptance of evidence.

3) Automatic information retrieval for further support

If the data is structured and semantically annotated it is possible to write powerful decision support helpers. For example, in the demining context, a suggestion for which demining tool to use can be based on properties like hazard depth, soil type, metal level, slope, etc. It is simply required that the demining tool IOs in the database have a list of capability attributes. A query can then list all the tools that match these criteria in the system. Fig. 1 shows how such a query result might look. It partially shows a Hazard Area page with an example of how decision support can be used to help with planning a demining operation. The "Suitable tools for demining" is based on the operation depth (>15 cm), slope sensitivity (there is no slope in this HA) and soil type (Gravel). The system retrieves all tools that satisfies the criteria for this specific condition. Any tool that operates too shallow or cannot work in gravel is excluded. Every piece is automatically retrieved without any interference from the user. All data shown is "mock-up" data.


Suitable tools for demining (Based on soil type: Gravel a recommended: 15 cm)

	Operation depth	Sensitive to slope?	m² per day
MineWolf MW370	35	Yes	30000
Personnel with tool	30	No	10

Health care [edit]

Distance to closest health care facilities: Al Aadeisse hospital (0.75 km);

Surrounding cities/villages/locations [edit]

Aadaisse (0.74 km); Marjeyoun refugee camp A (0.78 km); Markaba (2.58 km); Rab

Figure 1: Decision support example.

F. Summary

The process from gathering information to taking a decision on actions is an iterative process in the presented concept. Each step in the workflow is depicted in Fig. 2. First step is to gather information about the IOs (mainly infrastructures, settlements and evidence), from the acquired information impact and probability is calculated. The classification is done based on previous information. The last step is to analyze and prioritize between areas for possible actions or further information gathering. After this step the process starts all-over again.

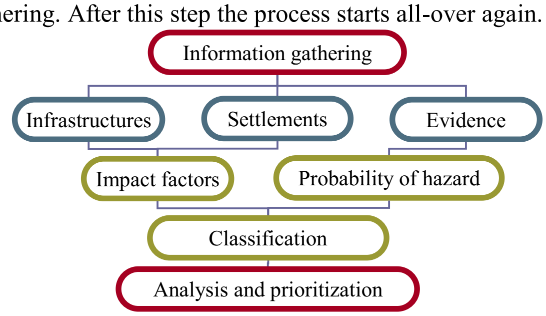


Figure 2: A schematic view of the information gathering and analysis process in the demining domain.

With the gathered information a classification could for example be: *Confirmed hazard area with high impact on the surrounding society* or *Suspected hazard area with negligible impact on surrounding society*.

G. Limitations

A set of unsolved problems have been identified that needs to be addressed in the future:

- **The association problem**. When there are several observations (evidence) for the same indicator they may be counted several times.
- **Age of the evidence**. Some evidence gets less valid with age. This is not considered in the model.
- **Evidence probability uncertainty** is not considered in the Bayesian equations.

IV. REALIZATION

In this section the realization of the concept described in III is presented.

A. Implementation

The following implementation of HAIMS is solely focused on the demining domain. However, we argue that the concept can be applied to other hazardous contamination tasks. The developed system for the described model in section *III* is based on a semantic extension to MediaWiki [14]. Therefore the implementation goes under the name *MineWiki* as the name describes these two components. The Mediawiki [15] allows for different extensions and libraries to MineWiki that are required to meet the functionalities defined in section *III*. Furthermore, similar to [5] it can work as a collaborative workspace for deminers to discover, manage and store relevant information described in [16]. Other advantageous features are:

- **Web based**. A user can log in from all over the world and use the system via any web browser with internet access.
- **Networked multiuser system**. Several users can use the same information and make changes simultaneously.
- **User right**. Users can have different reading and writing rights allowing for hidden and/or restricted information, user spaces, and different roles within an operation.
- **Distributed databases**. If the information or traffic becomes massive it is possible to have the database distributed on several servers.
- **Open ended usage**. The wiki style allows for creation of any kind of information. Everything from pages on refugee movements and the political situation in the country to mine defusing and standard operational procedures can be entered easily.
- **Traceability**. Every change is logged and can be traced over unlimited iterations. It helps to make sure all users are aware that they are responsible for what they are entering. This therefore improves the data quality.
- **Interaction support**. This involves multiple languages and the dynamic display of information (pages, IOs).
- **Built in API**. This allows for automatic retrieval or recording of information from other software and/or machines.

Semantic MediaWiki was selected due to ease of use and above listed built in functionality; however, the model is not restricted to this. The described concept can be implemented in any environment that allows semantic functionality.

This approach using semantic techniques also benefits strongly from having a Graphical User Interface (GUI). As a GUI to display information the MineWiki uses so called *(Web) pages*. A page can be used to display one or many IOs, or parts of IOs. Pages can also contain other information created by a user. The main page interface is a map with self-explanatory icons where each icon represents IOs and can be clicked to access more information. The system uses a connection to Google Maps^{TM} to support the overview for tactical decision makers. The map part of the GUI can be seen in Fig. 3. Each icon represents an IO and the coloured polygons their respective area shapes. Each icon links to information which can accessed by the user. In this case parts of the information are displayed when clicking on the icon.

Figure 3: Screenshot of the map GUI.

Also text format, for example tables, free standing text, numbers etc., are ways of displaying information from the database on a page – including graphs, timelines or text. Data is also available via SQL-queries and SPARQL commands, but to improve usability it is not required in the tool environment.

The semantic extension has its own script language which allows making *templates* that display information in a structured and dynamic way. Templates are used to make (and hide) the queries so that the user does not need to know anything about writing queries or scripts. E.g., the page template for the Hazard Area will not only provide information from the IO itself but the displayed information will dynamically change as other IOs changes in the system. The semantic system will automatically provide further information such as surrounding medical facilities, nearby contamination (with attributes like hazard type, age, distance etc.), nearby accidents (and other evidence of hazards), nearby important infrastructure and settlements.

As discussed before the data gathering for the calculation of impact and probability is significantly simplified with the semantic approach. The system automatically calculates the relevant information within a certain radius and makes the semantic calculations required – e.g. impact factors and probability values discussed earlier. The user is presented with the total value and if requested by the user, will the tool present the list of individual factors and their value contribution.

Forms are used to simplify the input. This will direct the user when entering information in the system. The form makes sure the data entered is in the right format and type (discussed in section *III.D – Information Quality*). The forms will also provide decision support as suggestions of information to enter.

The tool also supports task assignment such as Request for Information (RFI) which further supports collaboration.

B. Validation

We have validated the tool by demonstrating it to experts in the demining community. A scenario for a demining operation was performed in a field demonstration. The experts had between few and 20 years of experience from various countries. The general conclusion was that the presented solution seemed very relevant to the demining community.

V. CONCLUSION

The concept described in this paper has several individual ideas to improve the IM of hazard information:

- A first step standardizes the evidence recording (evidence forms).
- A novel way of using of the evidence to do hazard probability calculations.
- A novel way to calculate impact and to gather impact factors.
- Using semantic techniques to do decision support and calculations, including those mentioned above.

It is claimed that the concept enables advantageous anticipatory and responsive planning. By using a reasoned and tested systematic process it highlights gaps in knowledge and information. This prompts further searches for information to acquire all available relevant data. The concept also establishes a common and accepted approach to the process of gathering information. The uniformed approach supports the quality and efficiency of planning by managing information on hazard contamination, especially for demining operations. The product of the planning and decision process is the definition of hazard areas that are minimized with respect to the balance between risk, societal impact and available information and to the national land release criteria of the country. We have argued that these suggested ideas will improve the treatment of information in the realm of identification and prioritization of hazard elements. The presented information model and implementation has been demonstrated to end-users in the field of demining which have acknowledged its usefulness.

Note: the tool is a stand-alone system and works as presented but integration with already existing information management systems for deminers (e.g. IMSMA) is desirable to unlock the full potential of the concept as the concept builds on previously known/gathered information.

ACKNOWLEDGMENT

The work presented in this paper was financed via the D-BOX collaborative project, EU FP7 grant agreement number 284996. The work is based on background work done by D-BOX partners. In particular the authors want to mention the requirements analysis work done by Astrium (Franco Curatella, Brian Howmans), important ideas, warnings and other feedback from demining end users and stakeholders primarily represented by BACTEC (Stewart Grainger).

REFERENCES

[1] O. Merrouche "The Long Term Educational Cost of War: Evidence from Landmine Contamination in Cambodia" The Journal of Development, 47:3, 399-416 2011.

[2] J. P. Trevelyan, "Technology Needs for Humanitarian Demining." Department of Mechanical and Materials Engineering, the University of Western Australia 2000.

[3] B. Van de Walle, M. Turoff, & S. R. Hiltz, (Eds.). (2009). Information systems for emergency management. ME Sharpe.

[4] P. Eachus and B. Short. Decision Support System for Intelligence Analysts. In Intelligence and Security Informatics Conference (EISIC), 2011 European (pp. 291-291). IEEE.

[5] M. C. Madison, A. J. Cowell, R. S. Butner, K. Fligg, A. W. Piatt, L. R. McGrath, and Ellis, P. C. Knowledge encapsulation framework for technosocial predictive modeling. Security Informatics, 1(1), 1-18, 2012.

[6] Shafer, G., 1976. "A mathematical theory of evidence" Princeton University press, Princeton, New Jersey.

[7] T. Denoeux, "Analysis of evidence-theoretic decision rules for pattern classification" Pattern recognition, 30, 7, 1095-1107, 1997.

[8] IMSMA, http://mwiki.gichd.org, accessed 2015.

[9] P. Lacroix, H. Santiago, and N. Ray, "MASCOT: Multi-criteria Analytical SCOring Tool for ArcGIS Desktop." International Journal of Information Technology and Decision Making, 13(6):1135-1159, 2014.

[10] A. Larsson, L. Ekenberg, A. Wessel, and H. Bach, "Non-technical Survey: A model for evidence –based Assessment" The journal of ERW and mine action" 14.1, 2010.

[11] R.. Guha, R. McCool, and E. Miller, "Semantic search." In Proceedings of the 12th international conference on World Wide Web (pp. 700-709). ACM, 2003.

[12] M. Bajic, A. Krtalic, C. Matic, and D. Vuletic, "Minefield Indicators and Analytical Mine Contamination Assessment in Scientific Projects and in Practice", International Symposium Humanitarian Demining, 2009.

[13] U. B. Kjaerulff and A. L. Madsen, "Bayesian Networks and Influence Diagrams", Springer 2008.

[14] M. Krötzsch, D. Vrandečić, and M. Völkel, Semantic mediawiki. In The Semantic Web-ISWC 2006 (pp. 935-942). Springer Berlin Heidelberg, 2006.

[15] D. J. Barrett, MediaWiki. " O'Reilly Media, Inc.", 2008.

[16] F. Curatella et al. "Toward a Multifaceted Platform for Humanitarian Demining" 13th IARP Workshop on Humanitarian Demining and Risky Intervention, 2015.