

Homeland Security Information Representation: A Case for Standardization

HARRY D. TUNNELL IV Indiana University School of Informatics,
Indianapolis
hadtunne@iupui.edu

ABSTRACT

Advances in digital technology have improved educator access to information about their particular discipline. Homeland security is such a new discipline, however, that it does not have a large body of organized, domain-specific, public digital material. Consequently, educators and students will have to rely on resources from other specialties until the homeland security discipline matures enough to develop adequate discipline-specific, publically available digital sources. Fortunately, information from other domains can have a great deal of relevance. Unfortunately, the material is often not represented in a manner that reflects a relation to homeland security — which makes it a challenge to search and retrieve. The development of standards to represent homeland security data regardless of where it is stored can improve the ability to access digital materials, irrespective of discipline, that complement homeland security education and training programs. One near term solution is to develop a metadata standard that incorporates homeland security specific business metadata, which will allow the accurate description of any relevant digital artifact regardless of where it resides.

INTRODUCTION

Homeland security is a new discipline in the U.S. that gained popularity as a scholarly endeavor after the terrorist attacks of September 11, 2001 (Recca, 2012). More than 300 certificate and degree programs have been established in less than a decade (Ramirez & Rioux, 2012). Because of this newcomer status, publically available digital archives devoted to homeland security scholarly research are limited. Thus, homeland security researchers must often consult other types of databases in order to support their investigations. A solution to this problem is to develop standards that describe a homeland security digital artifact so that it is easily retrieved regardless of where it resides.

The advent of the public Internet and advances in digital technology allow members of academic, professional, and other communities to digitize materials about their particular discipline and make them accessible to others. Homeland security, a new discipline, does not have a strong tradition of scholarly research. This means that the digital resources that other domains maintain, such as digital libraries, databases, and search tools, are lacking in the homeland security sphere.

For example, the defense community has been able to depend upon the Defense Technical Information Center (DTIC) for more than six decades. The center maintains DTIC Online, an online resource of technical and other forms of defense related research. This is in stark contrast to the Naval Postgraduate School's Center for Homeland Defense and Security (CHDS), which has only been in operation since 2002 and maintains the Homeland Security Digital Library (HSDL). As the body of homeland security scholarly research grows, it will become increasingly important to represent digitally formatted homeland security information so that people can successfully retrieve it using modern information systems.

Even though the repository of homeland security information for education and scholarly research remains immature, personnel who lead academic and training programs have the option of turning to domains that do have a mature tradition of research. Investigating a highly technical subject (e.g., an adaptive wearable computer interface for homeland security use) is often difficult when relying exclusively on the homeland security literature. Conversely, one can examine technical topics using sources outside of the homeland security domain with the knowledge that a homeland security perspective is likely missing.

The availability of homeland security digital materials for research. The limited number of homeland security specific archives does not mean information is absent. Department of Defense (DoD) research, for example, can be a source of information for a homeland security researcher. Some military information is germane because there is overlap between the Department of Homeland Security (DHS) and DoD missions. Homeland defense is a DoD task that is complementary to homeland security and requires integration with law enforcement, emergency response, and other elements of DHS (Department of Defense [DoD], 2011). The DoD homeland defense mission provides an obvious linkage between the departments.

Linkages between homeland security and other fields of study are frequently less obvious, however. The adaptive wearable computer interface example is useful to highlight a potential relation between resources in other domains and homeland security research. Military operations occur in stressful high-risk environments and there is a body of scientific and applied research that is available about stress and the military environment (Cox & Fu, 2005; Harris & Hancock, 2003; National Research Council, 1997). Homeland security operations also occur in stressful high-risk environments, but it would take investigators decades to replicate this level reservoir of knowledge with homeland security conditions. Several studies already show that stress during military operations affects decision making and problem solving to such a degree that 70% to 85% of catastrophic mishaps are due to human error (Rash, Russo, Letowski & Schmeisser, 2009). Similar statistics are not available in the homeland security literature. Nonetheless, a researcher can infer from the earlier research that a homeland

security professional experiencing stress will have a degraded performance with an adaptive wearable computer interface.

For added factors regarding a computer interface, scientific research in the computing and psychology literature may prove useful even if context, scenario, and participant are unrelated to homeland security. Examples of possible digital archives to investigate for computing information are the Association for Computing Machinery (ACM) Digital Library and the Institute of Electrical and Electronics Engineers (IEEE) *Xplore* Digital Library. A database such as PsycINFO, which includes more than a century of research regarding human behavior, might also prove a valuable resource. All of these digital resources can inform homeland security researchers about the relation between computer interfaces, stress, and attention even though much of the research predates the establishment of DHS.

Current efforts to develop a repository of homeland security specific research such as HSDL are in their nascent stages. Consequently, digital materials from other domains are likely to be relied upon in homeland security education for the time being. The nature of the different sources can provide homeland security educators with opportunities to teach students how to evaluate the efficacy of other sources. In the case of adaptive wearable computer interfaces, numerous sources offer information about human factors, ACM and IEEE about technology, and PsycINFO about human behavior. Educators can insist that students evaluate how the differences in context, scenario, and participant might enhance or detract from an exploration of a homeland security professional's experience with the interface. Clearly, there are pertinent reservoirs of knowledge available to homeland security researchers. The limitation is that the information within them is normally not identified as related to homeland security.

Challenges retrieving homeland security digital artifacts. Even though there is the potential to glean useful information from related domains, this does not mean it is without challenge. Exploratory research about the relation between stress and attention — to understand how well a person experiencing stress can pay attention during task performance — provides an example of the difficulty people can face sifting through databases that are not organized for homeland security research. In the area of stress and attention seven data sources were examined and over 9,600 entries were returned (Table 1 shows the results from the top two searches). The high number of results indicates that there is a substantial body of research related to stress and attention even though all entries were not individually evaluated.

Table 1: A Sample of the Stress and Attention Search Strategy.

Database	Date	Search Strategy	Database Organization	Number of Articles
PsycINFO	June 2, 2011	Keywords: Stress AND Attention; 2000–2011; English; Young Adult; Peer- Reviewed Journal; Quantitative; Human	Abstracts	384
DTIC Online	June 25, 2011	Keywords: Stress AND Attention	Search terms were highlighted, abstracts were not displayed	8,680

The large number of items from DTIC Online is misleading. DTIC Online, in this instance, is not as effective as the PsycINFO database because DTIC Online displayed an overwhelming number of differently structured results (i.e., some with and some without abstracts or indexing [subject] terms). Part of the problem may be that DTIC Online employs at least three different models to describe the information in its databases. The Standard Form 298, Report Documentation Page, accompanies documents that can be printed. Nonprintable and multimedia items are submitted with a DTIC Form 530. Another form of electronic representation includes full abstracts.

A substantial amount of the information within DTIC Online predates the “Internet Age” and the establishment of DHS. Because of an artifact’s production date or the dissimilar methods of representation, the terms used to catalog an item might be inconsistent with current or emerging trends in digital Information Representation and Retrieval (IRR). A resulting dilemma is that being too specific with homeland security terminology can miss applicable material but being too general results in an overwhelming quantity of results (as in the case of stress and attention). Such representation problems are not unique to DTIC Online. The integration of older materials into a digital archive makes the IRR methodology a challenge. IRR problems are often a consequence of digitizing a large body of older multimedia information, particularly during periods of technology transition (Balk & Conteh, 2011; Deng, Wu, Lu & Brown, 2010; Nagy, 2007).

While the efforts to develop homeland security digital archives are ongoing, it is important to realize that creating vast publically available digital libraries is expensive and time consuming. It could take decades to establish extensive homeland security digital collections. Even digitizing small sets in an era of increasing budget constraints and personnel cutbacks may prove too costly. A 2008 effort that digitized 320 dissertations for a medical school repository cost USD \$23,562 and took 906 hours (Piorun, 2008). Rather than wait for the

development of adequate homeland security collections or rely upon digital artifacts that are not optimized for homeland security research, a new approach should be pursued. It is appropriate to develop standards to improve the IRR of current, and future, homeland security data regardless of the database. The lessons about using more than one model to represent information are instructive and similar problems should be avoided regarding homeland security information.

The central problem of any type of IRR in the digital space is to get the right information at the right time, regardless of other variables in the IRR environment (Chu, 2007). Potential solutions that can improve information access include establishing protocols to ensure that homeland security relevant information is properly represented so that it can be found with retrieval techniques common for new media (e.g., the Internet). This can be accomplished by designating clear standards (in indexes, abstracts, metadata, etc.).

The complementary part to representation is the retrieval itself. An information retrieval system should be able to find information that is described properly. The effort to digitize vast amounts of historical multimedia materials or create new IRR systems is time consuming, labor intensive, and costly (Lai, Fujinaga & Leive, 2005; Yacoub, Burns, Faraboschi, Ortega, Peiro & Saxena, 2005). Rather than create a new system from the ground up, it may be more practical and cost effective to standardize descriptions of homeland security digital resources. This will ensure that the best materials are available to researchers regardless of the digital archive in which they are located.

RECOMMENDATION

DHS has already established collaborations with several groups including DoD, IEEE, and many others. DHS should take advantage of these existing collaborations to improve homeland security IRR. DoD is an obvious partner because it already sponsors a homeland security digital library at the Naval Postgraduate School and it maintains other archives that contain related materials. This type of cooperation would correct some of the deficiencies identified in the 9/11 Commission report. The commission identified a lack of collaboration between organizations with a role in homeland security as a major shortfall and their recommendations included collaborative information technology (IT) solutions (National Commission on Terrorist Attacks Upon the United States, 2004).

The emphasis on collaborative IT means topics such as usability, human factors, interaction design, and other human-computer interaction topics are increasingly important to homeland security operations. IRR is an element in all of these topics. In order to tackle these problems, IRR collaboration must extend beyond DoD and DHS. Fortunately, numerous entities eminent in the IT domain such as IEEE are already partners with DHS (IEEE sponsors an international “Technologies for Homeland Security Conference”). Organizations such as IEEE can make important contributions because of their long-standing leadership in IT research and

standardization. These contemporary associations mean that a framework for collaboration on the homeland security IRR problem already exists.

Next, one must ask, “What IRR standards would improve the availability of materials for homeland security researchers?” Metadata for homeland security is an area to consider for improvement and standardization. Metadata is data about data; it is a method to describe information that is networked and in a digital format (Chu, 2007). The concept of metadata to aid retrieval is already proven, with broadly recognized standards such as the Dublin Core, digital object identifier, and Metadata Object Description Schema.

The metadata solution for homeland security information does not require a completely new standard. Business metadata provides the context and meaning of data that is represented by a computer and this is the type of metadata important to homeland security researchers (Inmon, O'Neil, & Fryman, 2008). Inadequate representation of a digital artifact's contextual information limits a researcher's knowledge of the artifact's primary content and business metadata mitigates a representation deficit (Winget, 2007). This approach is different from the technical form of metadata commonly used to design and maintain a specific application (Inmon et al., 2008; Stamm, McCleave, & Jagodnik, 2006). Unique metadata schemas in other domains have incorporated characteristics from previously established standards (Lai et al., 2005). This tactic will work for homeland security metadata as well. A standard such as Dublin Core can be used for the generic resource description and business metadata for homeland security specific attributes. Moreover, Dublin Core assimilation ensures that homeland security resource descriptions work with evolving technologies such as the Semantic Web, because Dublin Core is already used with a framework that describes resources on the Semantic Web (W3C, 2004).

Standardized homeland security business metadata promotes the integration of scholarly and practitioner information as well. The scholar-practitioner integration occurs because there is one benchmark for data collected across the various subdomains of homeland security (education, law enforcement, emergency management, etc.). Having one metadata standard means that a scholar or practitioner only has to learn one way to understand homeland security data in order to perform searches *regardless* of location and media/information type (e.g., video surveillance, PhD dissertation, or suspicious activity report).

Even though full implementation of metadata has a long way to go, the concept is well-established enough that an exclusive method to describe homeland security information in the databases of other disciplines is appropriate to consider. Moreover, if the homeland and national security communities establish a coordinated form of metadata they can guarantee the unique requirements of government documentation, such as security classification management, are adequately addressed. The standard can ensure, for instance, that a PowerPoint briefing about terrorists prepared in a U.S. Army Tactical Operations Center in

Afghanistan has metadata that allows it to be retrieved during collaboration between homeland defense and homeland security organizations facing a similar terrorism problem elsewhere. Integrated metadata standards can also support complementary data (e.g., cultural, regional, or social) from nongovernment databases, thereby promoting easier retrieval. Any standardization effort must consider the retrieval methods that are currently available, those expected to be in use in the future such as Semantic Web technology, and the user experiences of the researchers who are likely to search the materials.

CONCLUSION

The emphasis on collaboration between elements in and out of the homeland security community portends substantial information exchanges that will be stored in various databases (Department of Homeland Security, 2008). The complexity of retrieving material will increase due to the diversity of organizations that will manage homeland security related information. In addition to the obvious advantage for homeland security educators, improved IRR aids people who are outside of the direct homeland security domain but wish to develop or improve partnerships due to professional, scholarly, or commercial interests (Department of Homeland Security, 2008; Indiana Office of Energy and Defense, 2007). Therefore, efforts to standardize information representation have benefits that extend beyond the classroom.

Metadata is only one aspect of digitization and IRR (Nagy, 2007). However, it provides an important method to improve the retrieval of homeland security research immediately. Once metadata is agreed upon in the homeland security community, the standard can be promulgated among other communities of interest. It is important to consider standardization of homeland security digital artifacts now because the amount of homeland security scholarly research is relatively small. As a result, the volume of material is manageable and allows a standard to be more consistently applied since the metadata will be utilized as materials are produced. The application of a new standard after a large body of work exists is a costly problem that can be avoided by adopting unambiguous homeland security metadata now.

REFERENCES

- Balk, H. & Conteh, A. (2011). *IMPACT: Centre of competence in text digitisation*. Paper presented at the Proceedings of the 2011 Workshop on Historical Document Imaging and Processing: Beijing, China.
- Chu, H. (2007). *Information representation and retrieval in the digital age*. Information Today, Inc.: Medford, NJ.
- Cox, C., & Fu, D. (2005, November). *AI for automated combatants in a training application*. Paper presented at the IE 2005: Proceedings of the second Australasian conference on Interactive Entertainment: Sydney, Australia.
- Deng, F., Wu, Z., Lu, Z., & Brown, M. S. (2010). *BinarizationShop: A user-assisted software suite for converting old documents to black-and-white*.

- Paper presented at the Proceedings of the 10th Annual Joint Conference on Digital Libraries: Gold Coast, Queensland, Australia.
- Department of Defense. (2011). *Joint Publication 1-02 Department of Defense dictionary of military and associated terms*. U.S. Government: Washington, D.C.
- Department of Homeland Security (2008). *Department of Homeland Security information sharing strategy*. U.S. Government: Washington, D.C.
- Harris, W. C., & Hancock, P. A. (2003, October). *Field assessment of cognitive performance under stress*. Paper presented at the Proceedings of the Human Factors and Ergonomics Society 47th Annual Meeting: Denver, CO.
- Indiana Office of Energy and Defense Development. (2007). *Indiana defense asset study and BRAC-affected communities analysis*. Indianapolis, IN.
- Inmon, W. H., O'Neil, B., & Fryman, L. (2008). *Business metadata: Capturing enterprise knowledge*. Morgan Kaufmann: Boston, MA.
- Lai, C., Fujinaga, I., & Leive, C. A. (2005). *The challenges in developing digital collections of phonograph records*. Paper presented at the proceedings of the 5th ACM/IEEE-CS Joint Conference on Digital Libraries: Denver, CO.
- Nagy, G. (2007). *Digitizing, coding, annotating, disseminating, and preserving documents*. Paper presented at the proceeding of the 2006 International Workshop on Research Issues in Digital Libraries: Kolkata, India.
- National Commission on Terrorist Attacks Upon the United States. (2004). *The 9/11 Commission report*. Government Printing Office: Washington, D.C.
- National Research Council. (1997). *Tactical display for soldiers: Human factors considerations*. National Academy Press: Washington, D.C.
- Piorun, M. (2008). Digitizing dissertations for an institutional repository: A process and cost analysis. *Journal of the Medical Library Association*, 96(3), 223–229. doi: 10.3163/1536-5050.96.3.008
- Ramirez, C. D., & Rioux, G. A. (2012). Advancing curricula development for homeland security education through a survey of DHS personnel. *Journal of Homeland Security Education*, 1(1), 6–25.
- Rash, C. E., Russo, M. B., Letowski, T. R. & Schmeisser, E. T. (Eds.). (2009). *Helmet-mounted displays: Sensation, perception and cognition issues*. U.S Army Aeromedical Research Laboratory: Fort Rucker, AL.
- Recca, S. (2012). Homeland security education: Reading the tea leaves. *Journal of Homeland Security Education*, 1(2), 2–5.
- Stamm, R. J., McCleese, Y., & Jagodnik, A. J. (2006, March). *A network-centric approach to enhanced national airspace security*. Paper presented at the 2006 IEEE Aerospace Conference: Big Sky, MT.
- W3C. (2004). RDF primer: W3C recommendation 10 February 2004. Retrieved from <http://www.w3.org/TR/2004/REC-rdf-primer-20040210/#dublincore>
- Winget, M. (2007). *Digitizing & providing access to contextual cultural materials: The liner notes digitization project*. Paper presented at the proceedings of the 7th ACM/IEEE-CS Joint Conference on Digital Libraries: Vancouver, BC, Canada.
- Yacoub, S., Burns, J., Faraboschi, P., Ortega, D., Peiro, J. A., & Saxena, V. (2005). *Document digitization lifecycle for complex magazine collection*. Paper presented at the proceedings of the 2005 ACM Symposium on Document Engineering: Bristol, UK.