from: *Proceedings of the Eighth International Conference on Semantic Technologies for Intelligence, Defense, and Security*, Fairfax, VA, (STIDS 2013), CEUR, vol. 1097, 33-40.

IAO-Intel

An Ontology of Information Artifacts in the Intelligence Domain

Barry Smith
University at Buffalo
NY, USA

Tatiana Malyuta
CUNY, NY, USA
Data Tactics, McLean, VA

Ron Rudnicki
CUBRC, Buffalo
NY, USA

William Mandrick
Data Tactics
McLean, VA, USA

David Salmen
Data Tactics
McLean, VA, USA

Peter Morosoff
E-Maps, Inc.
Washington, DC, USA

Danielle K. Duff
I2WD
Aberdeen, MD, USA

James Schoening
I2WD
Aberdeen, MD, USA

Kesny Parent
I2WD
Aberdeen, MD, USA

***Abstract*—We describe on-going work on IAO-Intel, an information artifact ontology developed as part of a suite of ontologies designed to support the needs of the US Army intelligence community within the framework of the Distributed Common Ground System (DCGS-A). IAO-Intel provides a controlled, structured vocabulary for the consistent formulation of metadata about documents, images, emails and other carriers of information. It will provide a resource for uniform explication of the terms used in multiple existing military dictionaries, thesauri and metadata registries, thereby enhancing the degree to which the content formulated with their aid will be available to computational reasoning.**

Keywords—ontology; information artifacts; military doctrine; intelligence analysis; interoperability; data services environment

I. BACKGROUND

Standardization of terminology has been important from the very beginning of organized warfare. Imagine the Chinese trying to pass reports down the Great Wall using fire beacons without standardization of the signals used. In the Revolutionary War, General Washington directed Friedrich Wilhelm von Steuben to write the drill manual for the Continental Army [1] so that all units would use and respond uniformly to the same commands.

In our own era, DoD has directed development and use of the *DoD Dictionary of Military and Associated Terms* (Joint Publication 1-02) as the paramount terminological standard for military operations [2]. JP 1-02 helps to enable joint warfare by (a) advancing consistency in communications and (b) facilitating consistent interpretation of commands. Military dictionaries and related terminology artifacts continue to be developed, addressing these and a series of additional aims, including: (c) compiling lessons learned (outcomes assessment); (d) providing controlled vocabularies for official reporting; and (e) enhancing discoverability and analysis of data.

Such artifacts have until recently been conceived by analogy with traditional free-text dictionaries published in forms designed to maximize utility to human beings. Most existing doctrinal and related lexica and thesauri not only provide little aid to computation, they also suffer from the fact that multiple such resources have been (and continue to be) developed independently, in divergent and often non-principled ways. The result is that identical data may be classified and described entirely differently by different agencies, and the consequences of the resultant failures of integration (for example in the case of registries of persons of interest) are all too familiar. Increasingly, however, it is recognized that there is the need for a unified approach to description and classification of information resources (see for example [3], [4]), and the DoD has recognized at an official level that, to advance discoverability and analysis in the age of Big (military) Data, new approaches are needed that can enable computational retrieval, integration and processing of data. Thus Directive 8320.02 [5], the latest version of which is dated August 5, 2013, requires all authoritative DoD data sources to be registered in the DoD Data Services Environment (DSE) [6]. It further requires that all salient metadata be discoverable, searchable, retrievable, and understandable:

> Data, information, and IT services will be considered understandable when authorized users are able to consume them and when users can readily determine how those assets may be used for specific needs. Data standards and specifications that require associated semantic and structural metadata, including vocabularies, taxonomies, and ontologies, will be published in the DSE, or in a registry that is federated with the DSE.

We shall return to the DSE below. First, we present our own strategy for realizing these important goals.

II. THE INFORMATION ARTIFACT ONTOLOGY

The Information Artifact Ontology (IAO) was originally conceived in 2008 as part of an effort to master the Big Data accumulating in the wake of the Human Genome Project in the context of biological research [7]. Its goal was to aid the consistent description of biological data emanating from multiple heterogeneous sources. The goal of IAO-Intel is analogous: it is to provide common resources for the consistent description of information artifacts of relevance to the intelligence community in a way that will allow discovery, integration and analysis of intelligence data from both official and non-official sources.

When biomedical informaticians work with databases, publications and records generated by experimental research or medical care they focus primarily on what these artifacts *describe* (for example on the genes or proteins which form the subject matters of a given journal publication, or on the symptoms or diseases reported in a given clinical note).

Similarly, when intelligence analysts work with source data artifacts, then they, too, focus primarily on what the data in these artifacts describe, for example on the military units whose movements are recorded in a given shipping report, or on the vulnerabilities of a given forward operations base as described in some force protection assessment.

But while the primary focus concerns in both cases the topic or subject of the artifacts in question, both also require a secondary focus, targeted to the artifacts themselves, through which information about these topics is conveyed. Such artifacts have attributes – including format, purpose, evidence, provenance, operational relevance, security markings – data concerning which (often called 'metadata') is vital to the effective exploitation of the reports, images, or signals documents with which the analyst has to deal.

The dichotomy between *focus on entities in the world* and *focus on the information artifacts in which these entities are represented* is fundamental to the work reported here. IAO relates precisely to the objects of this secondary focus. An information artifact (IA), as we conceive it, is an entity that has been created through some deliberate act or acts by one or more human beings, and which endures through time, potentially in multiple (for example digital or printed) copies. IAO thus deals with information in the forms it takes when it has been deliberately fixed in some medium in such a way as to become accessible to multiple subjects. Examples are: a diagram on a sheet of paper, a video file, a map on a computer monitor, an article in a newspaper, a message on a network, the output of some querying process in a computer memory.

III. Goal of IAO-Intel

The goal of IAO-Intel is to support the effective handling of data concerning those attributes of IAs that are relevant to the purposes of intelligence analysis. To describe such attributes coherently we need to distinguish:

– the *particular* information artifact of interest, tied to some particular physical information bearer: the photographic image on this piece of paper retrieved from this enemy combatant; the email created by this particular author on this specific laptop; the target list compiled for this particular artillery unit on this particular date;

– the *copyable* information content that is carried by the artifact in question. The photographic image may be printed out in multiple paper copies; the email or target list may be transmitted to multiple further recipients. The information content that is copied or transmitted thereby remains in each case one and the same.

IAO-Intel provides ontology terms relating both to official documents and to non-official (source) artifacts. It provides also a set of relations to be used when we wish to represent the fact that, say, IA #12345 *is-about* some given person, or *uses-symbols-from* some specified symbology, or *links-to* some second IA #56789, and so forth,

IAO-Intel is designed from the start to provide the needed supplement in a way that will create semantic interoperability of data retrieved from different types of sources through an incremental process of semantic enhancement as described in [8], [9] and [10]. It is designed to allow automatic retrieval of all documents in a given collection of heterogeneous sources which involve a particular creator, or a particular type of intelligence report, or a particular type of weblink, or have been declassified under the authority of a particular agency, or are operative within a given time window.

Importantly, IAO-Intel is not designed to *replace* existing doctrinal or other standards created to guide human beings or computer applications in the creation and description of documents in accordance with defined formats or document architectures. Rather, its purpose is to allow the results of using such standards to generate the needed metadata in a uniform, non-redundant and algorithmically processable fashion. Moreover, the broad scope of IAO-Intel means that the metadata generated in relation to *official* documents will be of a piece with the metadata incrementally accumulating in relation to all information artifacts of relevance to the IC – the metadata will consist, in every case, of annotations to IAs formulated in ontology terms drawn not only from IAO-Intel but from the entire suite of DSGS-A ontology modules.

Thus while using existing standards for human or computer-aided creation or description of IAs does indeed allow us to retrieve data pertaining to IAs prepared in accordance with these standards, for IAs of other sorts the existing approach will fail. Only an ontology-based approach along the lines here proposed can, we believe, demonstrate the sort of flexibility and consistent expandability which are needed in today's dynamic and data-rich environments.

IV. Explication and Annotation

Currently a draft version of IAO-Intel is being applied within the framework of the US Army's Distributed Common Ground System (DCGS-A) Standard Cloud (DSC) initiative as part of a strategy for the horizontal integration of warfighter intelligence data [9]. Two sorts of application are currently being used to enable the ontology to support computer-aided retrieval and analytics. First, is *explication* of general terms used in source intelligence artifacts and in data models, terminologies and doctrinal publications which provide typologies of intelligence-related IAs. Second, is the *annotation* of the instance-level information captured by such IAs.

Explication is performed by providing definitions of such *general* terms using the resources of IAO-Intel and of the domain ontologies (such as Agent or Event ontologies) being developed within the DSGS-A framework. Annotation is performed by associating ontology terms with data about *particular* persons, events, or places in given information artifacts.

Table 1. Sample Types and Subtypes of Information Artifacts

IAO	IAO-Intel (examples)
Report	Intelligence Report (FM 6-99.2, 126)
Summary	Electronic Warfare Mission Summary (FM 6-99.2, 87)
Diagram	Network Analysis Diagram (from JP 2-01.3, II-51)
Overlay	Combined Information Overlay (JP 2-01.3, II 33)
Assess-ment	Assessment of Impact of Damage (FM 6-99.2, 53)

Estimate	Adversary Course of Action Estimate
List	List of High-Value Targets (JP 2-01.3, II 61)
Order	Airspace Control Order (FM 6-99.2, 17)
Matrix	Target Value Matrix (JP 2-01.3, II-63)
Template	Ground and Air Adversary Template (JP 2-01.3, II-57)

The goal of explication is to ensure that the data captured in annotations is semantically enhanced in a way that enables computational integration and reasoning along the lines described in [11], [12]. The goal of annotation is to aid retrieval of information about specific persons, groups, events, documents, images, and so forth, where this information is conveyed through source documents using disjointed and disparate systems for designation.

V. Strategy for Building IAO-Intel

Our strategy for building IAO-Intel is to extend the draft IAO to include terms and definitions tailored for the intelligence domain and specifically for the needs of our DSGS-A ontology initiative. The strategy has the following parts.

First, IAO-Intel is created by downward population from the draft IAO reference ontology. That is, the highest level terms of IAO-Intel are defined as specializations of terms from IAO along the lines illustrated in Table 1. The coverage domain of IAO-Intel will be determined incrementally on the basis of requests from analysts and other SME communities and through incorporation of terms from doctrinal publications and relevant high-level data models and document classifications.

Second, we use these sources to identify the dimensions of attributes along which IAs will be annotated. The selected dimensions are constructed in such a way as to be *orthogonal* in the sense in which, for example, color is orthogonal to shape – thus ontology branches built to represent different dimensions of attributes will contain no terms in common. This will enable these branches to be structured following the principle of single inheritance (thus as true hierarchies) [13].

Third, we create *low-level ontology modules* (LLOs) corresponding to each of these orthogonal dimensions. LLOs are small single-dimension attribute lists or shallow hierarchies designed to advance ease of maintenance and surveyability of the ontology and to provide a growing set of simple component terms which can be used:

1. to *construct more complex terms*, both terms for inclusion in IAO-Intel, and terms to be used to generate inferred classifications in application ontologies created for specific local purposes, along the lines described in [10];
2. to *define* the terms of the IAO-Intel ontology and of its sister ontologies within the DSGS-A framework;
3. to *explicate* the meanings of terms standardly used by different agencies, or by different groups of SMEs, or by different existing and future systems to describe such artifacts in a logically consistent way that is designed to allow integration of data and enhanced analytics;
4. to *annotate* instance data pertaining to particular information artifacts used by the intelligence community – for instance analysts' reports; harvested emails; signals data; and so forth.

The goal is that IAO-Intel should support integration of data annotated using different standard terminology resources. To bring this about, the constituent terms of such resources will be explicated using terms from IAO-Intel so that the artificial composite terms used in certain official terminologies and exchange model resources (along the lines of 'VehicleInspectionJurisdictionAuthorityText') will be broken down logically into constituent elements. This will provide a means to avoid the combinatoric explosion that is threatened by traditional approaches. Some composite expressions – for example 'Essential Element of Friendly Information (EEFI)' – will indeed be included in pre-composed form in the IAO-Intel ontology, but only where they are either defined in doctrine or already established as part of relevant SME vocabularies.

The modeling task for which compounds such as 'VehicleInspectionJurisdictionAuthorityText' were designed is addressed in our framework by allowing single data entries to be annotated by multiple ontology terms (sometimes linked by appropriate relations). A record in one of the tables containing data about an IED can be annotated, for example, both with 'IED Event' (based on its aboutness) and with 'EEFI' (based on its importance). A particular plan for the Intelligence Preparation of the Battlefield can be annotated as being at the same time a Plan (based on its purpose), a Government Document (based on its source), a Report on Air Defenses (based on its aboutness). It can be annotated also through relations, for example through *located-at* linking the source of the plan to some city or building and linking the planned air defenses to some region of interest.

Currently, military terminology resources generally fail to follow established best practice principles for the formulation of definitions. For example, they often confuse terms referring to components of information artifacts with terms referring to the entities in reality which those information artifacts are about. The "WTI Improvised Explosive Device" Glossary, for example, defines Method of Emplacement as:

> The description of where the [improvised explosive] device was delivered, used or employed.

Similarly the DCGS-A Logical Data Model defines Cover-Concealment as:

> information about geographical features that provide protection from attack or observation.

Use of IAO-Intel in tandem with corresponding domain ontologies allows us to explicate CoverConcealment (properly so-called) as:

> a geographic feature which *has-role* CoverRole,

and to explicate CoverConcealmentInformation as:

> IA which *is-about* CoverConcealment,

where CoverRole is defined as:

> the Role acquired by a given geographic feature when it is used to provide protection from attack or observation.

VI. Maintaining and Evaluating IAO-Intel

To maintain the IAO-Intel term collection over time we will create feedback links to enable users of the ontology to request new terms and to report errors. We are also working on an objective validation process which will enable us to determine how requested terms should be treated, distinguishing options such as: 1. incorporation into IAO-Intel or into some associated reference ontology, 2. incorporation into an application ontology maintained for some local purpose, 3. being marked as a synonym of some existing ontology term.

We are identifying, and where necessary constructing *de novo*, the domain ontologies that will need to be used in the definition of complex terms, and defining the relations that will link IAO-Intel terms with terms in these domain ontologies. These ontologies, too, will be extended over time on the basis of input from users.

We are also testing a series of objective criteria to be used in evaluation of IAO-Intel and other DCGS-A ontologies, starting with simple numerical measures of (a) term requests received and dealt with, and (b) uses of terms in *definitions*, *explications* and *annotations*. IAO-Intel will allow us to keep track of the number of information artifacts that make reference to individuals falling under a given class, and these metrics too can be used to assess the relative importance of this class within the ontology framework taken as a whole. While not definitive, such measures will help guide our judgments concerning the content and structure both of IAO-Intel and of its associated domain ontologies.

VII. Organization of IAO-Intel

Given the importance of the dichotomy between primary (topic) and secondary (artifact) focus, a central role in IAO-Intel is played by what we call

– Information Content Entities (ICEs) are *about* something in reality (they have this something *as a subject*; they *represent*, or *mention* or *describe* this something; they *inform us about* this something). Aboutness may be identifiable from different perspectives. Thus one analyst may interpret a given ICE as being about the geography of a given encampment; another may view it as providing information about the morale of those encamped there.

All major classes of information artifacts involve ICEs – simply because all major classes of information artifacts are about something. A plan of action, for example, is *about* a certain group of persons and goals and the types and ordering of actions that will be used to realize these goals. Even a document that has been written in code will be assumed by an analyst to be about something (for what, otherwise, would be the reason for its creation?). Typically, an information artifact such as a copy of a newspaper will be associated with multiple ICEs at successive levels of granularity, including separate articles within the newspaper, separate sentences within these articles, and so on.

In addition to ICEs, we distinguish also:

– *Information Bearing Entity* (IBE). An IBE is a material entity that has been created to serve as a bearer of information. IBEs are either (1) self-sufficient material wholes, or (2) proper material parts of such wholes. Examples under (1) are: a hard drive, a paper printout (e.g., a report); and under (2): a specific sector on a hard drive, a single page of a paper printout.

– *Information Quality Entity* (IQE). An IQE is the pattern on an IBE in virtue of which it is a bearer of some information.

– *Information Structure Entity* (ISE). An ISE is a structural part of an ICE; speaking metaphorically, it is an ICE with the content removed: for example an empty cell in a spreadsheet; a blank Microsoft Word file. ISEs thus capture part of what is involved when we talk about the 'format' of an IA.

The term 'information artifact' can now be used to refer either 1. to some combination of ICEs and ISEs (roughly: the IA as body of copyable information content); or 2. to some concretization of ICEs and ISEs in some IBE in which some IQE inheres (the information artifact is: this content here and now, on this specific computer screen or this printed page). Different information artifact types will differ in different ways along these dimensions, as illustrated in Table 2.

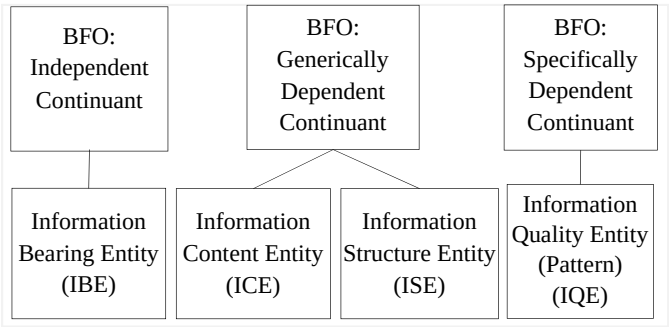


Figure 1. Continuants in the IAO framework

VIII. IAO and the Basic Formal Ontology

Figure 1 shows how IAO and IAO-Intel are being built to conform to Basic Formal Ontology (BFO), the upper-level architecture used in the DSGS-A ontologies [14]. IBEs are, in BFO terms, *independent continuants* (they are entities made of physical matter). An IBE is a physical entity that is created or modified to serve as bearer of certain patterned arrangements – for example of ink or other chemicals, of electromagnetic excitations. An IQE is a quality of an IBE which exists in virtue of such patterned arrangements and which is interpretable as an ICE or ISE. Such an IQE is created when some physical artifact is deliberately created or modified to support it (patterned to serve as its bearer). IQEs are BFO:*specifically dependent continuants* (SDCs) – entities which require some specific physical bearer but which are not themselves physical. Each IBE and IQE is restricted at any given time to some specific location in space. (If you display the same digital image twice on your desktop, then there are two IQEs on your desktop, which are – at some level of granularity – indistinguishable copies of each other.

ICEs and ISEs, in contrast, are what BFO calls *generically dependent continuants* or GDCs. This means that they are entities – such as a pdf file or an email – which can be copied from one physical bearer to another and thus may exist simultaneously in multiple different IQEs, which are called

'concretizations' of the corresponding GDC. Each GDC is concretized by at least one specific IQE inhering for example in the tiny piles of ink on the piece of paper in your pocket or in differentially excited pixels on your screen. When the GDC is copied, then a new IQE is created on a new physical information bearer, as when a new pattern of characters is created on the screen of the recipient of an email. This second pattern is a copy of the pattern created on the screen of the sender. The GDC itself exists simultaneously both at its original site and at the site to which it has been transmitted. GDCs can thus be multiply located.

BFO *relations* between ICEs, ISEs, IQEs and IBEs can be set forth as follows:

ICE *generically-depends-on* IBE
ISE *generically-depends-on* IBE
IQE *specifically-depends-on* IBE
ICE *concretized-by* IQE
ISE *concretized-by* IQE

IAO contains in addition relations which allow us to formulate metadata concerning attributes of IAs such as author, creation date, classification status, and so forth, and to annotate also components of IAs such as the To- and FromAddress components of email headers. The ToAddress of email message *m*, for example, is defined as:

> a collection of at least one email addresses of the intended recipients of *m*, each with at most one optionally associated name.

The set of relations can be extended to include also relations involving documents, document parts and document collections, such as *retrieved-from*, *curated by*, and so forth.

When we consider examples such as those provided in Table 2, then it becomes clear that, when IAO-Intel is applied to the explication of terms involved in describing instance-data relating to real-world IAs, then multiple artifacts may need to be distinguished. Consider, for example, a pdf file stored on some specific laptop. When we address what is meant by the (copyable) content of this file, then we recognize that this content may be copied in multiple ways, for example: to a pdf file using the same version of the Acrobat software and on the same operating system, to a pdf file using a different version of the Acrobat software, using characters from the same or a different character set, by being printed out on a piece of paper, and so on. The annotation of instance data with information of this sort may be important for example in investigating the provenance of given information artifacts which lie at the end of long chains of copying and processing involving multiple authors and computer systems. One potential application of IAO-Intel is to the systematic annotation of data pertaining to such chains.

Matters are complicated further when we go deeper into the question of how IAs are stored inside the computer. Given a generically dependent continuant which is the pdf file stored in the hard drive on some given laptop, there is a specifically dependent IQE which is (roughly) the pattern of 1s and 0s in the magnetic coating of the hard drive. When the entirety of this pdf file is displayed on your screen, then there is a further specifically dependent IQE which is the corresponding pattern of pixels on your screen. Both of these IQEs are concretizations of a corresponding GDC.

Note that we do not assume that all portions of IAO-Intel will be of equal utility in applications for the IC. We do, however, believe that to achieve clarity of explication in the treatment of source data artifacts will require clear definitions of the upper-level terms in the IAO, and a clear understanding of the relations between them.

TABLE 2: DIMENSIONS OF INFORMATION ARTIFACTS (IAS)

Information Artifact	**IBE**	**ISE**	**ICE**
MS Word file (.doc, .docx)	Hard drive (magnetized sector)	MS Word format	Varies
XML file	Hard drive (magnetized sector)	XML V 2.0 format	Varies
MS Excel 2010 file (.xls, .xlsx)	Hard drive (magnetized sector)	MS Excel 2010 format	Varies
KML file	Hard drive (magnetized sector)	KML	Map overlay
JPEG file (.jpg)	Hard drive (magnetized sector)	JPEG format	Image
Email file (with embedded attachments	Hard drive (magnetized sector)	Internet Message Format (e.g., RFC 5322 compliant)	Message
USMTF Message file	A specific government network	USMTF Format	Message
Passport	Paper document; (may include photographs, RFID tags)	ID formats, security marking formats …	Name, Personal data, Passport number, Visas …
Title Deed	Official paper document	Varies	Varies
Report	Varies	Varies	Varies
Overlay Sheet (e.g. Map Overlay Sheet – see Figure 2)	Acetate sheet	MIL-STD-2525 Symbols; FM 101-1-5 Operational Terms and Graphics	Map overlay

IX. ATTRIBUTES OF INFORMATION ARTIFACTS

Information artifacts have attributes along a number of distinct dimensions, treated in LLO modules of the IAO. Terms in these modules will be applied to explicate information relating to IAs of different types, and to annotate data pertaining to IA instances with the help of relations mentioned above. Some dimensions of IA attributes are common to all areas, both military and non-military, including: *Purpose*, *Lifecycle Stage*

(draft, finished version, revision); *Language*, *Format*, *Provenance*, *Source* (person, organization), and so forth.

Along the dimension of *Purpose* we distinguish:

- Descriptive purpose: scientific paper, newspaper article, after-action report
- Prescriptive purpose: legal code, license, statement of rules of engagement
- Directive purpose (of specifying a plan or method for achieving something): instruction, manual, protocol
- Designative purpose: a registry of members of an organization, a phone book, a database linking proper names of persons with their social security numbers

whereby it should be stressed that one and the same IA may of course serve multiple purposes.

As is shown in Table 3 IAO-Intel will include additional LLOs relating to attributes of importance to the intelligence domain such as: *Classification*, *Encryption Status*, *Encryption Strength*, and so forth. IAO-Intel will also include terms representing specific IA *Purposes* such as: informing the commander, providing targeting support, intelligence preparation of the battlefield.

TABLE 3. DIMENSIONS OF INFORMATION ARTIFACT ATTRIBUTES

Role in the Intelligence Process (JP 3-0, III-11) Priority Intelligence Requirement (PIR) Commander's Critical Information Requirement (CCIR) Essential Element of Information (EEI) Essential Element of Friendly Information (EEFI)	
Confidence Level (JP 2.0, Appendix A)	
Highly Likely	Unlikely
Likely	Highly Unlikely
Even Chance	
Discipline (JP 2.0, I-5)	Intelligence
Legal	Signal
Ideology	Human
Religion	Rumor intelligence
Propaganda	Web intelligence
Intelligence Excellence (JP 2.0, II-6)	
Anticipatory	Complete
Timely	Relevant
Accurate	Objective
Usable	Available

Table 3 illustrates fragments of some of the dimensional hierarchies specific to IAO-Intel, with their doctrinal sources.

X. EXAMPLES OF USE OF IAO-INTEL IN ANNOTATION

As should by now be clear, IAO-Intel relates not merely to textual documents but to information artifacts of all types including maps, videos, photographic images, websites, databases, and so forth, both unstructured source documents and official documents of many different varieties. Consider, the Modified Combined Obstacle Overlay (MCOO), taken from JP 2-01.3 [15] and illustrated in Figure 2. (We refer to this as example IA#1 in what follows.) An MCOO is defined as:

> A joint intelligence preparation of the operational environment product used to portray the militarily significant aspects of the operational environment, such as obstacles restricting military movement, key geography, and military objectives.

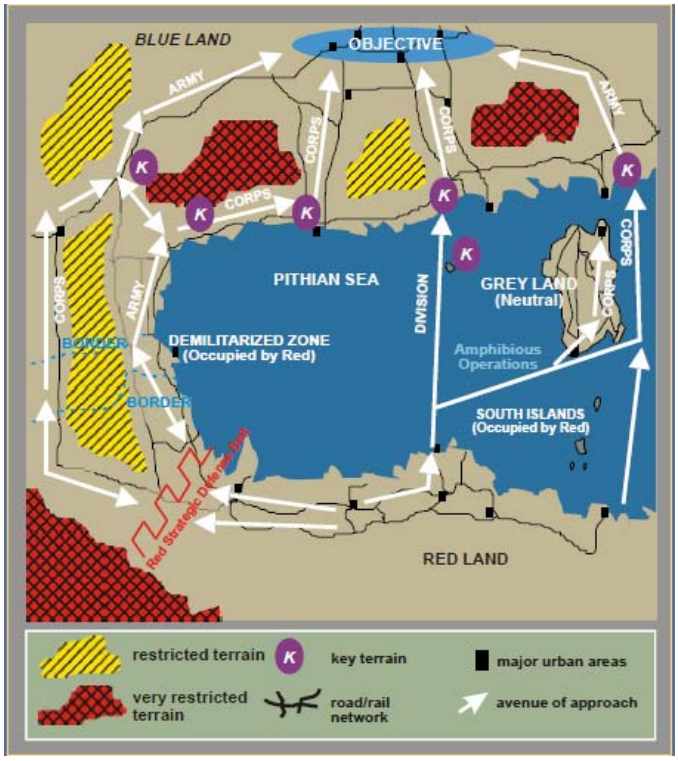


Figure 2: Modified Combined Obstacle Overlay (example IA#1)

We assume that IA#1 has been prepared as part of some given plan, IA#2. Both IAs #1 and #2 will then be referred to in multiple further IAs including multiple databases compiled during planning, execution and outcomes assessment. Relevant terms used in the data models associated with these data models will have been explicated using terms from IAO-Intel. The latter terms can then be used along the lines described in [9] to create annotations to both #1 and #2 on the basis of the fact that they are referred to in the databases in question. The results will include, for example:

a) annotations to the attributes of IA#1*:*

- ICE: MCOO
- IBE: Acetate Sheet
- *uses-symbology* MIL-STD-2525C
- *authored-by* person #4644
- *part-of* plan IA#2

b) annotations relating to the aboutness of IA#1

- Avenue of Approach
- Strategic Defense Belt
- Amphibious Operations
- Objective

and so forth. Used in conjunction with the skill ontology and the person database the annotations above will enable a planner

to retrieve (for example) all MCOOs relating to amphibious operations authored by persons with certain skills.

Consider, as a second example, a collection of documents prepared according to FM 6-99.2 [16], for example of types:

Intelligence Report [INTREP]
Intelligence Summary [INTSUM]
Logistics Situation Report [LOGSITREP]
Operations Summary [OPSUM]
Patrol Report [PATROLREP]
Reconnaissance Exploitation Report [RECCEXREP]
SAEDA Report [SAEDAREP]

Suppose further that we need to cross-reference these with comparable sets of documents prepared by other commands, and that we need to do this in such a way as to extract and process the information computationally. FM 6-99.2 provides definitions of the mentioned report types, but does not take the step of formulating these definitions computationally. IAO-Intel addresses this problem by providing a common, algorithmically useful, set of ontology terms that is designed to allow consistent explication of these and related types as they appear in different doctrinal resources. The results can then be used for computer-aided aggregation of the data represented using corresponding IA types, cross-checking of mismatches, and so forth.

XI. The DoD Data Services Environment

We can now return to Directive 8320.02 and address the relevance of the work reported above to its successful implementation. As we saw, the Directive requires that 'all salient metadata be discoverable, searchable, and retrievable' through use of the DoD Data Services Environment (DSE) [6]. DSE's numerous data sources include 35 'supporting taxonomies' derived from pre-existing terminology resources. Problems arise, however, because the latter have been constructed on the basis of multiple distinct methodologies (for example as concerns the formulation of definitions). When, on August 25, 2013, the DSE was queried for information on "location", the DSE reported 660 possibly relevant sources of information. When the DSE was queried for "unit types," 882 possibly relevant sources of information were reported. When types of "ground vehicles" were queried for, 175 possible relevant sources of information were reported. Such redundancies present obstacles to discovery, search and retrieval. They arise because different compilers of authoritative data describe entities of the same types in heterogeneous ways. This thwarts the sort of coherent integration that is required for the mounting of what, in [6], we referred to as the "massing of intelligence fires".

One problem is that while the terms in thesauri and glossaries can be used in annotations, the value derived therefrom is limited above all because they do not allow the benefits of inferencing and of rapid introduction and definition of new terms which are provided by a framework of well-constructed ontologies along the lines described in [10]. There we show how reference ontologies can be quickly expanded with new content to meet emerging data representation needs and in such a way that data annotated with the newly added terms is automatically integrated with existing data.

Imagine, for example that we have two large bodies of data describing (A) chemicals (properties, costs, manufacture, transport, supply, and so forth), and (B) explosives manufacture (raw materials, persons and skills involved, processes and equipment and safety measures used). We will have satisfied Directive 8322.20 in maximizing discoverability if we annotate each body of data in accordance with corresponding term repositories, which we can assume to have been independently developed. Suppose now, however, that we are called upon to integrate the data in (A) with the data in (B). Here these annotations will likely provide no assistance, which will in turn lead to calls for the creation of a third term repository to be used in efforts to annotate the combined (AB) data. The results of these efforts will then once again likely provide no assistance when (AB) data itself needs to be integrated with, say, data about explosives financing.

Where, in contrast, the systems for annotating (A) and (B) reflect a common ontological approach, then new annotation resources for the merged data can be easily be developed by reusing the initially developed ontologies in the formulation of both composite terms and corresponding definitions [10].

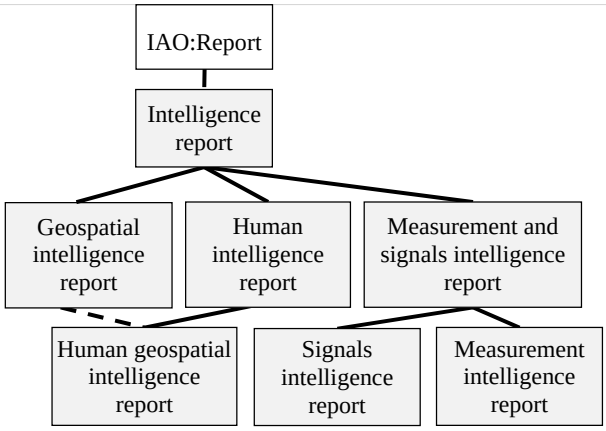
A further problem is that the need to create new terminology resources for the annotation of such merged content may lead to the need for corrections of the initial terminology resources. Such corrections may have expensive consequences: either they will break interoperability with the results of earlier annotation efforts, or – if resources are invested to correct already existing annotations to make them conform to the new usage – they will have unforeseen consequences for third parties who have been relying on the older resources to be maintained consistently through time. Such problems are minimized where terminology resources are developed in tandem from the very start as parts of a single suite of ontology modules developed using common principles, exactly as is proposed by our DSGS-A strategy. We believe that only a strategy of this sort can satisfy the requirement that data, information, and IT services are 'made visible, accessible, understandable, trusted, and interoperable throughout their lifecycles for all authorized users.' [5]

XII. Semantic Technology is Not Enough

The strategy underlying DSE has much in common with a strategy adopted widely in the semantic technology community under the heading of Linked Open Data, a strategy often involving the use of the Dublin Core Metadata Element Set as controlled vocabulary. We believe that the Dublin Core can serve as reliable controlled vocabulary for describing IA data only where the information artifacts in question are themselves artifacts formulated using RDF or some other W3C recommended syntax, and unfortunately this is not the case for many of the artifacts at issue here. We believe further that the Linked Data approaches cannot solve the problems of silo-formulation in the IC for the results outlined already in section XI above. The semantic technology community draws a distinction between two levels of interoperability: Level 1, resting on shared term definitions (for example drawn from the Dublin Core), and Level 2, of what is called Formal Semantic Interoperability. As is recognized at [17], Level 1 is 'so open-

ended that it quickly leads to a proliferation of custom-built solutions incompatible with each other, such as metadata expressed in document formats that require customized software to read and data models that cannot easily be mapped to generic, interoperable representations such as those expressed in RDF.' Level 2 is designed to solve these problems by requiring that all IAs are described via metadata formulated using RDF. Unfortunately RDF (or even OWL) is no panacea. Multiple conflicting ontologies can be formulated in RDF terms, yet still remain conflicting.

The solution, again, must rely on shared development of a single suite of modularized ontologies, in which not only the same formal language is used, but also consistent definitions populating downward from a common upper level such as BFO – and we note in this connection a parallel with the way in which joint doctrine is elaborated, in a process that is designed to ensure (at least ideally) that the same term is defined and used consistently across the 80 plus Joint Publications (JPs) that address the various aspects of joint warfare in accordance with JP 1-02 [2].



The above IAO-Intel terms are defined by using terms from the ontologies below with the help of relations such as *is-about*, *created-by*, *derives-from* and so forth [7].

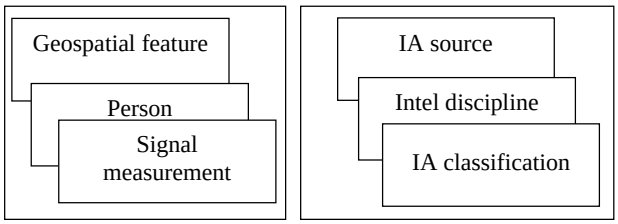


Figure 3. Top: Terms from IAO (unfilled) and IAO-Intel (grey) ontologies. Taxonomical hierarchies: asserted – solid lines, inferred – dashed lines. Bottom left: Domain ontologies. Bottom Right: IAO-Intel LLOs.

XIII. Conclusion

To summarize: IAO-Intel forms part of a collection of ontologies that is being applied primarily to the explication of data models and other terminology resources of importance to DCGS-A. The terms in these ontologies are linked together logically in virtue of the fact that each ontology uses terms which are defined in terms of other ontologies belonging to this same suite (as illustrated in Figure 3). This strategy for ontology development has been tested in use over several years in the domain of biomedical informatics, and is gradually being adopted also in other domains, including for example the domain of modeling and simulation, where the identifying authoritative data sources is needed to ensure realistic scenarios [18]. One principal feature of the strategy is that it provides a standard means for defining new ontologies in light of emerging needs, in a way that guarantees consistency with the ontologies already created and with the data annotated in their terms. We believe that this feature makes the strategy particularly useful in addressing the emerging challenges to the intelligence analyst in accordance with DoD directives concerning discovery, retrieval and search.

Acknowledgments

Work on IAO-Intel was supported by I2WD. Thanks are due also to Mathias Brochhausen, Werner Ceusters, Mélanie Courtot, Janna Hastings, James Malone, Bjoern Peters, Jonathan Rees, and Alan Ruttenberg for their work on IAO.

References

[1] Friedrich Wilhelm von Steuben, Regulations for the order and discipline of the troops of the United States, 1792, http://x.co/1dJEk.

[2] Department of Defense Dictionary of Military and Associated Terms, 2013, http://www.dtic.mil/doctrine/new_pubs/jp1_02.pdf.

[3] Leo Obrst, Patrick Cassidy, "The need for ontologies: Bridging the barriers of terminology and data structure", Geological Society of America Special Paper 482, 2011.

[4] Leo Obrst, Terry Janssen, Werner Ceusters (eds.), Ontologies and Semantic Technologies for the Intelligence Community. Amsterdam: IOS Press, 2010.

[5] Sharing Data, Information, and Information Technology (IT) Services in the Department of Defense, DoD Instruction 8320.02, August 5, 2013, http://www.dtic.mil/whs/directives/corres/pdf/832002p.pdf.

[6] DSE Data Services Environment, https://metadata.ces.mil/dse.

[7] https://code.google.com/p/information-artifact-ontology.

[8] David Salmen, Tatiana Malyuta, Alan Hansen, Shaun Cronen, Barry Smith, "Integration of intelligence data through Semantic Enhancement", Proceedings of the Conference on Semantic Technology in Intelligence, Defense and Security (STIDS), 2011, CEUR 808, pp. 6–13.

[9] Barry Smith, Tatiana Malyuta, David Salmen, William Mandrick, Kesny Parent, Shouvik Bardhan, Jamie Johnson, "Ontology for the Intelligence Analyst", CrossTalk: The Journal of Defense Software Engineering, November/December 2012, pp. 18–25.

[10] Barry Smith, Tatiana Malyuta, William S. Mandrick, Chia Fu, Kesny Parent, Milan Patel, "Horizontal integration of warfighter intelligence data. A shared semantic resource for the Intelligence Community", Proceedings of STIDS Conference, 2012 (CEUR 996), pp. 112–119.

[11] Ron Rudnicki, Werner Ceusters, Shahid Manzoor and Barry Smith, "What particulars are referred to in EHR data?", American Medical Informatics Association 2007 Annual Symposium, 2007, pp. 630–634.

[12] Ron Rudnicki, "DCGS-A Ontology Program Explication Procedures", MS, 2013.

[13] Barry Smith and Werner Ceusters, "Ontological Realism as a methodology for coordinated evolution of scientific ontologies", Applied Ontology, 5 (2010), pp. 139–188.

[14] Basic Formal Ontology 2.0, http://ontology.buffalo.edu/BFO/Reference.

[15] Joint Publication 2-01.3 Joint Intelligence Preparation of the Operational Environment, 16 June 2009.

[16] U.S. Army Report and Message Formats (FM 6-99.2), April 2007, http://armypubs.army.mil/doctrine/DR_pubs/dr_a/pdf/fm6_99x2.pdf.

[17] Dublin Core User Guide, Last modified September 6, 2011, http://wiki.dublincore.org/index.php/User_Guide.

[18] Saikouy Diallo, Jose Padilla, "Military Interoperability Challenges", Handbook on Real-World Applications in Modeling and Simulation, Wiley, 2012, pp. 298–332.