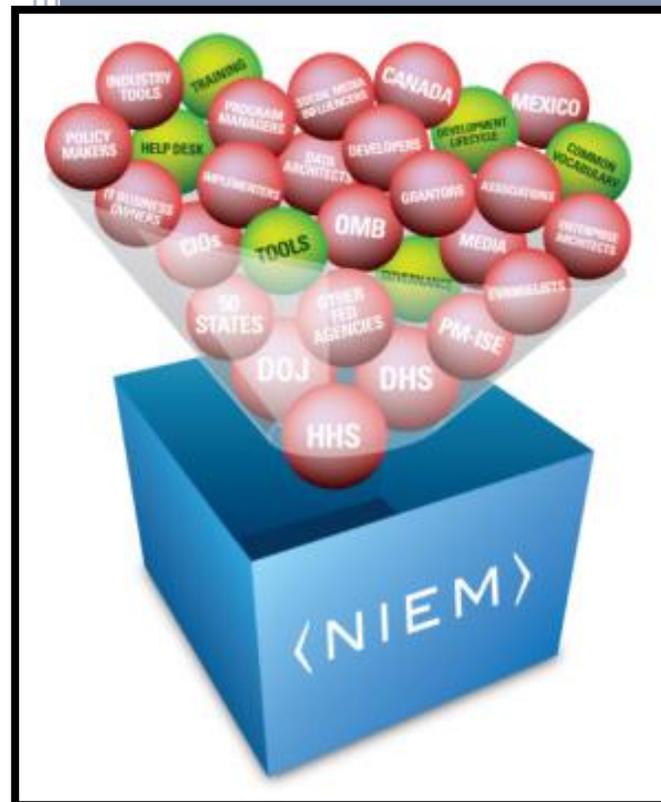




IJIS Institute

NATIONAL INFORMATION EXCHANGE MODEL 2.1 INTERNATIONALIZATION SUPPORT OVERVIEW



IJIS Institute

**IJIS Technical Advisory
Committee (I-TAC)**

March 2013

Principal Contributor

Joel Byford, Manager – Soos Creek
Consulting, LLC

This project was supported by **Grant No. 2009-DB-BX-K107** awarded by the *Bureau of Justice Assistance*. The Bureau of Justice Assistance is a component of the Office of Justice Programs, which also includes the Bureau of Justice Statistics, the National Institute of Justice, the Office of Juvenile Justice and Delinquency Prevention, and the Office for Victims of Crime. Points of view or opinions in this document are those of the author and do not represent the official position or policies of the United States Department of Justice.

CONTENTS

INTRODUCTION.....	1
Internationalization	1
Localization	1
NIEM 3.0	1
OVERVIEW.....	2
Date & Time Formats.....	2
Currency Formats	2
Multi-Lingual Data	3
Multi-Lingual Code Lists.....	3
Multi-Lingual XML Elements	4
CONCLUSION	5
ABOUT THE IJIS INSTITUTE.....	6
REFERENCE MATERIAL	6
Acronyms and Abbreviations	6

INTRODUCTION

The National Information Exchange Model (NIEM) was initially developed in the United States and has its original roots in criminal justice and public safety. Over time, it has evolved into a more general public sector data model and paradigm to facilitate information sharing across Federal, state, and local agencies.

This evolution and high degree of collaboration has been extremely successful and has not gone unnoticed by other countries. This attention brings with it the strong likelihood of international collaboration and further adoption of a standard data model for public sector information sharing worldwide.

These opportunities, while exciting, heighten our awareness for more global support within our data standards community. Many differences exist between various regions and countries worldwide and as such, any international data model will need to take these differences into account and provide a means for effectively supporting them going forward.

This whitepaper is intended to provide a synopsis of NIEM's current active approach to supporting internationalization and suggests some general software industry approaches to potentially improve upon the progressive work already completed. This paper will be of particular interest to NIEM Technical Architecture Committee (NTAC) and NIEM Business Architecture Committee (NBAC) members as they plan for changes in upcoming releases of the specification. It also serves as a baseline impact assessment for international communities wishing to determine how to leverage NIEM in their data standardization projects.

Internationalization

Before discussing various aspects of supporting a global data model, one must first become familiar with two key terms used throughout the software industry in reference to globalization support. The first term is *Internationalization (I18N)*, which refers to the means and level by which any given software or package *enables multi-national support*.

Localization

While I18N references the ability for software to support a multi-national clientele, *Localization (L10N)* refers to the *actual translation and other work* required supporting a *specific locale, region or country*.

NIEM 3.0

When this whitepaper was published, NIEM 3.0 was in its initial Alpha 1 form. At that time, none of the aspects of I18N or L10N discussed by this whitepaper had been addressed. With that said, the NTAC and NBAC have both expressed a strong commitment to globalization of NIEM and have suggested that some of the items listed in this paper may be addressed as NIEM 3.0 is finalized.

Internationalization (I18N)

- Ability to support multiple regions or locales.

Localization (L10N)

- Effort required to support a specific region or locale.

OVERVIEW

NIEM, from its inception, has been heavily based on international standards. It has also maintained a high-degree of flexibility in order to support as many different scenarios as possible. These provide the NIEM community with an excellent base to build upon, while embracing a more global clientele. Some of the most common aspects of I18N within NIEM are depicted in the following table and are briefly described in the text that follows.

I18N ASPECT	CHANGES REQUIRED FOR NIEM 2.1	SUMMARY
Date & Time Formats	<i>None</i>	As defined by the World Wide Web Consortium (W3C), Extensible Markup Language (XML) schemas all inherently support a common means of expressing dates and times with a built-in regional offset.
Currency Formats	<i>None</i>	Currently, ISO-4217 provides a comprehensive list of currencies supported by NIEM.
Multi-Lingual Data	<i>I18N Required</i>	Currently, any NIEM text field can include an ISO 639-3 language code however no locale code is currently included (e.g. UK English differs somewhat from US English).
Multi-Lingual Code-Lists	<i>I18N Required L10N Required</i>	Heavy usage of XML Substitution Groups in NIEM enable any number of code lists to be used, however there is no way to link like-codes with one another across code lists (e.g. Blue = Blau = Azul).
Multi-Lingual XML Elements	<i>L10N Required (Optional)</i>	NIEM element names only expressed currently in US English (e.g. PersonName). L10N efforts would be required to translate these elements to different languages.

Date & Time Formats

NIEM is based upon the W3C XML Schema specification (XSD), which, in turn, means the data type for dates and times is standardized and prevents any confusion surrounding dates and times.

Date	<code><date>2008-01-01</date></code>
Date/Time - No Time-Zone	<code><dateTime>2008-01-01T12:00:00</dateTime></code>
Date/Time - With Time-Zone	<code><dateTime>2008-01-01T12:00:00-01:00</dateTime></code>

Optionally, if a regional time-zone offset is desired, it can be added to the data and expressed with a minus or plus sign along with hour and minute offset.

Currency Formats

A comprehensive list of currency codes is supported by NIEM Amount data types through an optional code-enforced XML attribute using the ISO-4217 code list. This commonly used code list is used throughout the industry and it supports both current and legacy currency types.

```
<nc:Fee>
  <nc:ObligationDueAmount nc:currencyCode="EUR">
    200.00
  </nc:ObligationDueAmount>
</nc:Fee>
```

Since NIEM follows XML Schema, the decimal delimiter options are governed by the W3C.

Multi-Lingual Data

The data payload itself being exchanged can be expressed in any number of languages. The XML files themselves are traditionally encoded in UTF-8, which supports both single- and multi-byte languages equally as well. In addition, all text-based elements in NIEM support the inclusion of a language code using the ISO 639-3 code list, as shown in the following example.

```
<nc:Person>
  <nc:PersonDescriptionText nc:languageCode="deu">
    Alter als der typische Mann.
  </nc:PersonDescriptionText>
</nc:Person>
```

While this enables some level of flexibility, it does unfortunately omit a regional or locale identifier. This additional locale in conjunction with the language is important as many different variations or dialects for a language exist. For example, English used in the United States differs somewhat from English used in Canada from both a spelling and grammatical standpoint. In order to properly understand and translate data being sent, both locale and language should be supported.

Traditionally, software developers have used a two-part localization code to denote language/locale combinations which are defined by RFC 1766 and include codes such as the following:

CODE	REGION/LANGUAGE
en-US	English (United States)
en-GB	English (United Kingdom)
es-MX	Spanish (Mexico)
es-ES	Spanish (Spain)

Multi-Lingual Code Lists

NIEM makes extensive use of Substitution Groups, which, in turn, enables multiple ways to represent any given category of data—for example, a list of driver license jurisdictions can currently be expressed in more than 10 different ways today.

Adding a code list in a different language would be relatively trivial given the current architecture of NIEM; however, there is no way currently defined to allow one to link similar code lists together.

For example, two code lists can be authored that both express

identical color codes, just in different languages, but without a way to link these two lists together along with their associated values there would be no way to easily translate the code lists between partners. If the OASIS Organization for the Advancement of Structured Information Standards (OASIS) Genericcode



Specification¹ is adopted in the future by NIEM, then this issue would be resolved as generic code supports linking code lists together.

Multi-Lingual XML Elements

While not strictly necessary, it may be preferable for NIEM schemas themselves to be expressed in a language other than US English. This way, schema authors and system developers in different regions can better read and understand the standards being developed and published by agencies.

NIEM is currently defined in only US English, which means a L10N effort would be required for each language wishing to be supported. This effort would likely include the following two tasks:

- 1) **Translation** - Translate NIEM Elements, Attributes and Data Types into a different language.
- 2) **Mapping** - Develop a crosswalk/map between the two or more languages.

Ideally, the mapping would be developed in a machine-readable technology such as XSLT to facilitate translation between languages.

In addition to, or as an alternative to multi-lingual elements, support could be provided for multi-lingual labels and descriptions for elements. Such labels and descriptions could allow a schema to be rendered in a localized language even if instance documents used English tags. This strategy is utilized in the Resource Description Framework (RDF) standards.

¹ <http://www.genericcode.org/>

CONCLUSION

NIEM's heritage of reliance on the open standards community has served the standard well in paving the way towards internationalization. To date, NIEM has leveraged all of the available internationalization expressed by the XML Schema standard and it must now look to some other authorities to further improve NIEM. As new versions of NIEM are planned and implemented, it is expected that the committees will make internationalization and localization a high priority and consider the information articulated by this paper.

We continue to be encouraged at the rapid adoption rate NIEM has seen both nationally and internationally, and fully expect that the committees involved in furthering the use of NIEM will continue their efforts to make NIEM a standard without borders.

ABOUT THE IJIS INSTITUTE

The IJIS Institute unites the private and public sectors to improve critical information sharing for those who provide public safety and administer justice in our communities. The IJIS Institute provides training, technology assistance, national scope issue management, and program management services to help government fully realize the power of information sharing.

Founded in 2001 as a 501(c)(3) nonprofit corporation with national headquarters on The George Washington University Virginia Science and Technology Campus in Ashburn, Virginia, the IJIS Institute has grown to nearly 200 member and affiliate companies across the United States.

The IJIS Institute does its valuable work through the contributions of its member companies. The IJIS Institute thanks the I-TAC for their work on this document.

The IJIS Institute also thanks the many companies who have joined as members that contribute to the work of the Institute and share in the commitment to improving justice, public safety, and homeland security information sharing.

REFERENCE MATERIAL

The IJIS Institute
<http://www.ijis.org>

Best Practices for XML Internationalization
<http://www.w3.org/TR/xml-i18n-bp/>

Official NIEM Website
<http://www.niem.gov/>

RFC 1766 Language/Locale Codes
<http://www.ietf.org/rfc/rfc1766.txt>

ISO-4217 Currency Codes
http://www.iso.org/iso/currency_codes_list-1

Acronyms and Abbreviations

ACRONYM OR ABBREVIATION	DEFINITION
I-TAC	IJIS Technical Advisory Committee
I18N	Internationalization
ISO	International Organization for Standardization
L10N	Localization
NIEM	National Information Exchange Model
OASIS	Organization for the Advancement of Structured Information Standards
RDF	Resource Description Framework
UTF	Unicode Transformation Format
W3C	World Wide Web Consortium
XML	Extensible Markup Language
XSD	XML Schema specification