EIP v1.0-compliant Prototype Implementation: Project Proposal

1. Opportunity
There is little debate that knowledge workers spend an excessive percentage of their time searching for and trying to evaluate information resources. The lowest published and anecdotal figure identified to date estimates the amount as greater than 40%. In addition, the search often concludes with little confidence that the best available information was acquired. Suboptimal decision making is one result, deciding to reacquire (and thus duplicate) lost information is another. With the proliferation of repositories and data volumes doubling every 1½-2 years, these problems are expected only to worsen.

The Energy Industry Metadata Standards Initiative is pursuing an opportunity to significantly increase the efficiency of discovering, evaluating, and accessing distributed information resources. The use of standardized, structured metadata as the means to capture this opportunity is central to the vision of the Initiative (figure right). The core deliverable of the Initiative is a new profile of the well established international metadata standard ISO 19115. This new profile – formally, the “Energy Industry Profile of ISO/DIS 19115-1 and ISO 19115-2 v1.0” – has been developed by Energistics’ Metadata Standards Work Group, with requirements input and feedback provided by members of the global energy community.

2. Business Case
The Metadata Standards Work Group is proposing to demonstrate the promise of the Energy Industry Profile (EIP) v1.0 by delivering a prototype network of four searchable metadata systems federated via the internet (See figure below). The primary deliverable of the project is one of these four searchable systems. The metadata catalog in this system will be EIP-compliant, and the system as a whole offered to the industry as a reference implementation for others to duplicate internally. This reference implementation will model the role that would be filled by a search system owned and managed by any organization consuming metadata to enhance discovery and evaluation of distributed resources. The
other three systems in the prototype network are already operational and owned by organizations interested in partnering to demonstrate a common vision. The other system catalogs are not EIP-compliant, but are compliant with metadata standards transformable to EIP v1.0. These systems model the role of organizations that have resources of value to others, and which are interested in making those resources available for discovery and evaluation by publishing their associated metadata.

This reference implementation of an EIP-compliant metadata server will allow Energistics to expand on its mission of providing open data exchange standards as a benefit to the upstream oil and gas industry. The deliverables of this project include the required documentation outlining the installation and configuration of an EIP-compliant metadata server which will allow interested organizations to create and populate their internal metadata in a standardized manner. This documentation will be freely-available on the Energistics website once the project is completed.

More specifically, eleven deliverables are proposed in Section 4, of which the first eight are considered critical to the core objectives of the project. These eight deliverables provide key supporting applications needed to enable robust implementations of such systems, the knowledge transfer to Energistics needed to maintain the reference implementation, and the workshops and webinar designed to encourage community adoption.

The resources required to execute the project will be provided by Energistics, Energistics’ Metadata Standards Work Group, Esri, the U. of Colorado/CIRES, and at their individual discretion, Sponsors and non-member Participants. Details are provided in Section 5. Resources at the U. of Colorado/CIRES will work under the guidance of Dr. Ted Habermann (NOAA/National Geophysical Data Center), who has accepted an invitation to participate in this project. Dr. Habermann is well known for his experience using ISO 19115 to document the metadata for large earth science data sets.
The project will commence following receipt of the minimum required funding, estimated at $125K and described in Section 6, and will proceed in five phases as outlined in Section 7.

As presented in Section 8, interested Energistics Members will be invited to sponsor the proposed project at a level between $10K and $25K commensurate with their Energistics membership fee. In addition to Members, selected non-member Participants, identified by their previous active engagement in the Industry Metadata Initiative, will be invited to contribute to the project. Participants that accept the invitation will be expected to contribute volunteer resources and metadata as well as the funding.

3. Approach

The prototype network will consist of four searchable metadata systems federated via the internet. Each system will contain a catalog of metadata records and will expose those records for search or harvesting using a standard catalog service interface (OGC CSW 2.0.2) that encodes the harvested records in XML format. This proposal seeks the funding needed to implement only the single metadata system to be operated by Energistics as a reference implementation for part of its open data exchange standards portfolio. The supporting tools needed by the community to create and validate EIP v1.0 XML metadata and to transform XML-encoded metadata in the other existing systems into EIP v1.0-compliant format will be developed with the prototype.

Open source software such as Esri Geoportal Server, GeoNetwork opensource, and Deegree offer the basic catalog-service capabilities that need to be provided by the Energistics system, including a metadata catalog, a web client to search for and inspect records in the catalog, and a service interface for search and harvesting the same records using the Open Geospatial Consortium (OGC) Catalog Service for the Web (CSW). Use of open source software and a standard catalog service interface leverages implementation of the prototype because existing software can be leveraged to enable operational capabilities at much lower cost that developing applications from scratch. This project will implement the Energistics system using Esri’s Geoportal Server software.

4. Deliverables (see Note 1)

In order of priority:

1. An online metadata catalog server enabled by EIP v1.0 using the Esri Geoportal Server open source software that will be operated by Energistics. The Geoportal Server metadata editor will be configured to create and update EIP v1.0-compliant XML metadata records. Also included is the knowledge transfer needed for Energistics to support and maintain the resulting system.


3. A software utility with associated Schematron rules that validates XML records as compliant with EIP v1.0.
4. Freely-available documentation on the Energistics website for the installation and configuration of an EIP-compliant metadata catalog server as an open standard.
5. Documented changes identified as improvements/fixes needed for EIP v1.1.
6. Two 1-day public workshops and at least one 4-hour webinar to enable and encourage community adoption, and solicit community input about potential future improvements. The two workshops will be held in Houston for community members capable of attending in person. The webinar will be scheduled to allow participation of those unable to attend either Houston workshop.
7. A web page (on energistics.org) that publishes EIP-specific code lists, and associated web service that validates terms as EIP v1.0-compliant code list values.
8. One or more metadata editor software applications, separate from that provided as part of the Energistics-owned Geoportal Server instance, capable of creating or updating EIP v1.0-compliant XML metadata records. Options include an ArcCatalog editor configuration (e.g., ANZLIC ISO Metadata Editor for 10), or other stand-alone application (CatMDEdit, Micka, etc. ...).
9. Additions to energistics.org web site:
   a) A registry of vocabulary terms that allows requests for i) all terms in a given vocabulary, ii) terms that are valid for a particular metadata element, and iii) will resolve concept identifiers to labels (English word) and definitions. Other features might include thesaurus-type relationships, and mappings between different vocabularies.
   b) A registry of catalogs that includes information in interfaces offered and metadata profiles supported.
10. Documentation outlining the design for a catalog implementation based on software other than Esri Geoportal Server (e.g., GeoNetwork opensource being used in European systems).
11. A .NET assembly and documentation to facilitate software development of applications capable of reading, manipulating, and writing EIP v1.0 XML metadata records.

5. Resources
The proposed project requires the resources enumerated below:
1. Leadership, guidance and coordination:
   - Leadership & coordination: Metadata Standards Work Group members
   - Guidance: Sponsor representatives (At Sponsor discretion)
2. Infrastructure (see Note 2):
   - Esri Geoportal Server software – Open source
   - Host server – Existing Energistics server in the iNetU environment
   - Web and Database server – Existing Energistics infrastructure in the iNetU environment
   - Server admin – Energistics (for Linux environment)
3. Software Development, Implementation, Training; Community workshops:
   From Esri:
   - Implement Energistics Geoportal Server, and effect the knowledge transfer necessary for its ongoing use and maintenance.
   - Develop supporting Geoportal applications and utilities.

5 June 2012
• Co-lead two community workshops (tentatively, June 1 & Sept 1) and one community webinar.

**From U. of Colorado/CIRES:**
• Develop the EIP v1.0 XML schema.
• Develop supporting XML transformation (XSLT) and validation (Schematron) software needed by the Energistics Geoportal for harvesting non-EIP metadata.
• Investigate feasibility of adapting NOAA metadata quality rubrics to EIP v1.0.
• Participate in two community workshops (tentatively, June 1 & Sept 1) and one community webinar.

4. **Metadata describing information resources.**

Sponsors will be invited, and Participants required to contribute metadata to Energistics to enable the project. The project will offer contributors a form-based mechanism to provide the metadata that does not require they know the EIP v1.0 standard. The metadata provided must have no constraints which prevent it being made publically accessible, though access to the associated information resources may be restricted.

6. **Cost**

Integrated project design and planning by all those contributing resources (Section 5) identified $125K as the cost to produce Deliverables 1-8 listed in Section 2 (above). The majority of these funds cover the cost of services provided by Esri and the U. of Colorado/CIRES to deliver:

• A Geoportal Server implementation functioning as part of the network described in Deliverable 1, as well as the supporting metadata editor and associated knowledge transfer.
• The supporting applications, utilities, and documentation described in Deliverables 2-4, 7-8.
• The workshops referenced in Deliverable 6.

The project will require two one-time service contracts, one each between Energistics and Esri ($98K), and between Energistics and U. of Colorado/CIRES ($25K) *(see Note 3)*. These contracts will result in no recurring costs since the system will be implemented in Energistics’ existing iNetU environment.

The final $2K of the budgeted project costs will be used to cover travel expenses for one member of the Metadata Standards Work Group to represent Energistics at the ISO/TC 211 34th Plenary and associated meetings, in particular, the TC 211 Project Team meeting that will initiate development of the new Technical Specification for the XML encoding of ISO 19115-1.

7. **Schedule**

Execution of the project will commence after funding has been secured, and the two supporting service contracts have been executed. The project will proceed in five phases:

1) Implementation of the Energistics reference implementation for an EIP-based Geoportal Server, and knowledge transfer training for Energistics.
2) Development/configuration of at least one EIP-compliant metadata editor, and one transform utility.
3) Delivery of the first community workshop, with the Metadata Standards Work Group.
4) Development of remaining supporting applications and utilities.
5) Delivery of the second community workshop, with the Metadata Standards Work Group.

8. Sponsors/Participants

Sponsors will be those Energistics members contributing funds to support the project. The project requests each Sponsor contribute at a level between $10K and $25K commensurate with their Energistics membership fee. Funding secured beyond $125K will be used both to produce the subset of Deliverables 9-11 (Section 4) that can be completed with the available funding, and to hold additional community webinars (Deliverable 6).

All Energistics members will be invited to become Sponsors. The following members have been actively engaged in the Industry Metadata Standards Initiative:

- BP
- Chevron
- ExxonMobil
- IHS Energy (through the SMT membership)
- Maersk
- Pioneer Resources
- Shell
- Schlumberger
- Statoil
- Total

In addition to Sponsors, the following non-member organizations, also previously active in the Industry Metadata Standards Initiative, will be invited to contribute as non-funding Participants in this project:

- USGIN
- USGS
- NOAA
- Apache
- CoreLogic
- Fugro Gravity & Magnetic Svcs
- Neftex

Notes

1. All artifacts produced by this project, including upgrades to the Geoportal Server package, will be placed in the public domain, available under an open source licenses to be determined by Energistics. The intention for this ready availability is to encourage and facilitate community adoption of the EIP.

2. Energistics’ iNetU environment provides two options:

   Preferred:
   Managed Linux Red Hat Gated Community Cloud Servers: 1 x vCPU's, 2GB RAM, 100GB SAN Disk Storage
   Included Features: VM Running on INetU Cloud w/ VMWare Enterprise Plus, Dynamic Resource Scheduler, vMotion
   Included Software: Linux Red Hat Enterprise x64, CLAM Antivirus.

   Alternative:
   Managed Windows 2008 Gated Community Cloud Servers: 1 x vCPU's, 4GB RAM, 100GB SAN Disk Storage
   Included Features: VM Running on INetU Cloud w/ VMWare Enterprise Plus, Dynamic Resource Scheduler, vMotion

3. Supporting proposals for contributions to be provided by ESRI and the U. of Colorado/CIRES are available upon request.