Grid-Enabling the Global Geodynamics Project: Automatic RDF Extraction from the ESML Data Description and Representation via GRDDL

L. Ian Lumb and Keith D. Aldridge
Earth & Space Science and Engineering
York University
Toronto, Ontario, Canada M3J 1P3
ian.lumb@rogers.com, keith@yorku.ca

Abstract

An eXtensible Markup Language (XML) based data model for the Global Geodynamics Project (GGP) has been previously developed. Mindful of the need to incorporate metadata into the description and representation, a Resource Description Framework (RDF) based approach is introduced that extends the previous data model. Specifically, use of RDF allows relationships to be described and represented, and will eventually result in an 'informal ontology'. The bottom-up approach makes use of GRDDL (Gleaning Resource Descriptions from Dialects of Languages) — a vehicle that allows for the extraction of RDF from XML according to rules. Because there exists some latitude in such extractions, complimentary top-down approaches will be required — especially when reconciling with formal ontologies. From this 'information science' perspective, GGP data has the potential to factor in the broader context being defined by the 'new geoinformatics'.

1. Overview

In a previous paper, Lumb & Aldridge [38] introduced an eXtensible Markup Language (XML) based data model for the Global Geodynamics Project (GGP). Building on this foundation (§2), and mindful of the need to incorporate metadata into the description and representation, a Resource Description Framework (RDF) based approach is emphasized here in §3. Key to this introduction from the bottom up is the recent arrival of GRDDL (Gleaning Resource Descriptions from Dialects of Languages) — a vehicle that extracts RDF from XML opposite rules. Because there exists some latitude in such extractions, a complimentary top down approach is suggested in §4. Knowledge-based investments such as this allow GGP data to play in broader contexts as highlighted in §5. Conclusions are drawn in the final section (§6).

2. ESML Representation of GGP Data

Recently, an XML-based data model was developed for the GGP [38]. Summarized schematically in Fig. 1, this model enables the transformation of semi-structured, discipline-specific (PRETERNA, Listing 1 of [38]) ASCII data into an XML-based representation (Earth Sciences Markup Language, ESML) via a converter. The converter makes use of rules that are detailed in a template (Listing 2 of [38]).

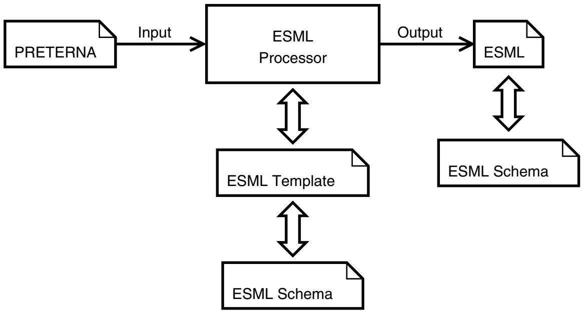

Figure 1. Simplified summary schematic of the ESML-based data model for the GGP (after [38]). Semi-structured, discipline-specific data in PRETERNA format are translated into an XML-based representation based on ESML via a translator. The translator operates according to rules defined in a template. The template relates the contents and structure of the ASCII-format file to the XML-based representation.

At the time of XML introduction, it was known that

ESML possesses a number of highly desireable characteristics [38] — i.e., it:

1. Makes use of XML Schema (XSD, [23])
2. Provides support for ASCII-format files
3. Has Earth Sciences affinities
4. Has industry standard affinities (i.e., DFDL, [1])

To this list, it is now possible to add that ESML is being adopted by large-scale, high-profile projects (e.g., [8]). An unanticipated benefit, however, was the ability to *leverage* ESML for the purpose of automatically generating metadata representations in RDF. Before turning attention to the RDF representation of GGP data in §3, it is necessary to more-completely describe the characteristics of GGP data.

Each GGP instrument generates the following three data files on a monthly basis [6]:

- Gravity and pressure data — The primary geophysical observables alluded to above, sampled on a regular basis, and identified by the `.GGP` filename extension
- Auxillary data — Geophysical observables (e.g., groundwater levels) that complement the primary observables, sampled on a regular basis, and identified by the `.AUX` filename extension
- Log data — Geophysical and other observables provided on an irregular basis, and identified by the `.LOG` filename extension basis

To fix ideas for the purpose of illustration, consider GGP data from September 1997 (`9709`) from station Strasbourg (`ST`) which has not required 'repair' (i.e., the data repair code is `00`.)[1] In this case, the primary, auxiliary and log files would have names `ST970900.GGP`, `ST970900.AUX` and `ST970900.LOG`. Because many GGP data acquisition systems remain DOS-based, use is still made of the legacy '8.3' DOS file naming convention; this lowest-common-denominator limitation factors directly in the metadata challenges reviewed below and identified originally in [38].

Lumb & Aldridge [38] suggested the introduction of an XML-based data model for the GGP through a monthly, two-step process:

- Represent and validate `ST970900.GGP`, `ST970900.AUX` and `ST970900.LOG` via ESML (per Fig. 1) — note that a separate template is required for each of these input files
- Aggregate the three ESML files (i.e., `ST970900_GGP.xml`, `ST970900_AUX.xml` and `ST970900_LOG.xml`) into a single file (i.e., `ST970900.xml`) via XSLT[2] or XInclude[3] (per Fig. 2)

It was also noted that annual aggregation could be achieved by an analogous transformation involving either XSLT or XInclude [38].

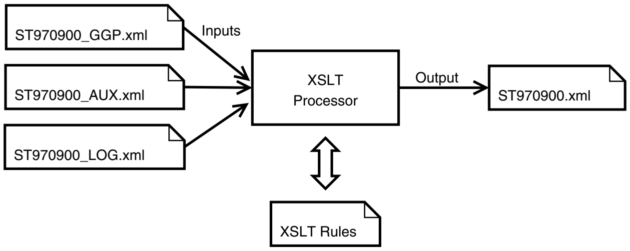

Figure 2. Simplified schematic of the ESML-based data model monthly summary for the GGP. Here the ESML-encoded input files are aggregated into a single monthly file via XSLT or XInclude (not shown). Note that all ESML files are encoded according to the ESML schema (not shown).

The introduction of a self-describing data model based on ESML offers many advantages. Perhaps most importantly from the scientist's perspective, such an approach directly targets one of their key motivators: The potential to easily automate the discovery of GGP data in time and/or space [38, 40]. This automation is revisited in §5. However, in and of itself, this XML-based approach *does not* address the challenges identified with respect to metadata [38] — i.e., data about data. Although specifics were not provided, an RDF-based representation was alluded to in [38] as a way of addressing these metadata-related challenges. In essence, this means converting per-station monthly ESML aggregates into RDF. The availability of automated schemes to make the ESML-to-RDF conversion clearly accelerates the ability to introduce metadata for the GGP, and is therefore the focus of the next section, §3.

With respect to metadata, Lumb & Aldridge [38] identified redundancies, and that:

> Application of ESML to GGP data also revealed that metadata is often ignored. The term 'latent

---

[1] GGP identifies a number of 'repair codes' that convey information on changes made to raw data — e.g., to account for gaps in the record. A compilation of repair codes is available elsewhere [6].

[2] eXtensible Stylesheet Language Transformations (XSLT, [24]) is a vehicle for structural (i.e., the conversion of one XML document type into another) and aesthetic (i.e., the formatting of one XML document type into another) transformation.

[3] XML Inclusions (XInclude) is a relatively new vehicle [21] for merging XML documents into a single, composite XML document.

> metadata' was introduced (§V of [38]) to describe this data about data that potentially exists, but is not presently evident or realized.[4] Although the examples presented suggest that there may exist tools to assist in extracting this metadata, such approaches are unable to overshadow extant challenges with metadata. As use of XML-based data models increases, and especially as use is made in large-scale projects (e.g., [9]), metadata challenges are receiving attention. In the long term, the most-attractive solution will necessarily have a semantic basis. This makes proactive investment in RDF, plus the emerging Semantic Web[5] and Semantic Grid[6] technologies, appear sensible.

The redundancies and latencies referred to above underline the simple fact that relationships have not been accounted for in descriptions of GGP data. Although RDF has the *potential* to make clear these relationships, Lumb & Aldridge [38] exercised cautious optimism:

> Longer-term investigations will be required to better understand the role of metadata in the context of the ESML-enabled GGP. An RDF-based approach also shows promise, but requires considerable investment to be realized.

It is the investment required to introduce RDF, therefore, that is considered in the next section.

3. Automated, Bottom-Up Extraction of RDF

Reconsidering the case of introducing an XML-based representation into the GGP, it's clear that the encoding *could* have been introduced manually — i.e., by manually interspersing ESML markup into a pre-existing ASCII-format file. Fortunately the use of a template, in tandem with an ESML processor, allows this encoding process to be automated. Somewhat fortuitously, automation is also receiving attention in the context of generating RDF representations, as this is regarded as key to accelerating progress with respect to the Semantic Web [27] — i.e., progress that serves to reduce the gap between the future potential versus the current reality of the Semantic Web [45, 29].

Already noted as staunch advocates on the importance of metadata [9, 42], the Marine Metadata Interoperability (MMI) Project has developed a tool that allows simple vocabularies to be transformed into ontologies[7] [26, 41]. Using their VOC2OWL 'editor', ASCII data in a tabular format can be transformed into a Web Ontology Language (OWL, [11, 29, 45]) representation. Thus VOC2OWL serves to automate the ASCII-OWL transformation from the 'bottom-up' — i.e., it uses a transformation process that starts at the lowest level of abstraction and proceeds towards higher levels. According to Bermudez & Graybeal [26], VOC2OWL works with vocabularies both flat (e.g., phone directories) and hierarchical (e.g., taxonomies). However, further investigation is required to determine VOC2OWL's suitability to observational data — e.g., such as that obtained in the GGP. Instead of proceeding with VOC2OWL, attention here focuses on GRDDL (Gleaning Resource Descriptions from Dialects of Languages, [5]) — a bottom-up mechanism for automating XML-to-RDF conversions. In the current context of the GGP, the intention is to transform ESML-represented GGP data into an RDF representation. While the outcome is not an ontology, *per se*, the resulting RDF representation certainly provides the *basis* for an 'informal ontology' [41].[8]

The GRDDL-mediated XML to RDF conversion is illustrated schematically in Fig. 3. An XML document and its corresponding schema are input to a GRDDL processor and result in RDF/XML on output [35]. The GRDDL processor is based on XSLT and carries out its transformation based on rules specified in one or more XSL files (not shown in this illustration). Because relationships are 'latent' (i.e., potentially exist, but are not presently evident or realized), GRDDL *extracts* RDF *from* XML (i.e., an XML document, its corresponding schema, and a set of XSL rules). Extraction of latent (i.e., pre-existing) relationships emphasizes the value of the 'objective' XML representation; it is quite different from an *introduction* in which the relationships are 'subjectively' detailed *in situ*. The double-ended block arrow at the top of Fig. 3 attempts to portray GRDDL in the broader context of the Semantic Web [20].

The GRDDL processor dereferences XML namespace Uniform Resource Identifiers (URIs, [10, 28]) to ob-

---

[4] Although Lumb & Aldridge [38] provided several examples of latent metadata, the recent availability of data [4] from the 26 December 2004 Sumatra-Andaman event [44] provides an even-more compelling illustration. In addition to `.GGP` data (i.e., 'standard' GGP Data) provided at a one-minute sampling interval, there is also the one-second data required for the analysis of shorter-period phenomena such as earthquakes and tsunamis. The introduction of this more-frequently sampled data necessitated the creation of a new filename extension, namely `.GGS`. The significantly greater volume of data also necessitates the use of compression utilities, and as a consequence these one-second files now have `.GGS.zip` or `.GGS.gz` extensions — reflecting use of ZIP or GNU ZIP formats, respectively. To reiterate the contention made by Lumb & Aldridge [38]: There is a wealth of latent metadata encapsulated in such filenames.

[5] Berners-Lee [28, pg. 122] defines the Semantic Web as a "... web of data that can be processed directly or indirectly by machines". Furthermore, this is a "... web of connections between different forms of data that allow a machine to do something it wasn't able to do directly" [28, pg. 185].

[6] The intersection of the Semantic Web and Grid Computing broadly defines the notion of a Semantic Grid; [34] and [48] provide deeper explorations in scientific contexts.

[7] Ontologies are defined in §4.

[8] Such machine-generated informal ontologies bear much in common with human-generated folksonomies — "... an emergent category that's defined from the bottom up" [43, pg. 138].

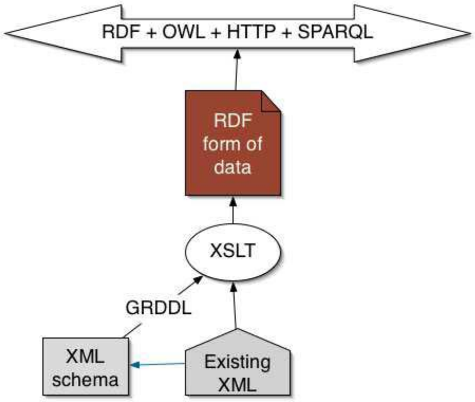

Figure 3. Schematic for the extraction of RDF from XML via GRDDL and XSLT ([14, Slide 22]). In this figure HTTP refers to Hypertext Transfer Protocol [7] while SPARQL is a query language for RDF [16].

tain the relevant rule-specifying XSL file(s) — in cases where the rules are not already known to it [35]. Given that URI dereferencing can be handled through various approaches, GRDDL also provides an explicit approach (see Listing 1). Thus the relevant transformation (e.g., `ggpEsml2Rdf.xsl` in Listing 1) is specified via the `data-view:transformation` attribute that has been appended to the XML root element. The resulting RDF/XML statements are taken to convey the 'intended meaning' of the input document.

Listing 1. Explicitly specified rules for GRDDL extraction of RDF from XML (after [35]). Note that the URL `http://www.ggp.org` is a ficitious one used here only for the purpose of illustration.

```
<xs:schema xmlns:xs=``http://www.w3.org/2001/XMLSchema↘
→''
    xmlns:data-view=``http://www.w3.org/2003/g/data-↘
    →view#''
    data-view:transformation=``http://www.ggp.org/2005/↘
    →g/ggpEsml2Rdf.xsl''>
```

Fig. 4 is a linear-systems-style schematic[9] that illustrates the use of GRDDL in the context of the GGP. In this figure, a month's worth of unrepaired GGP data (i.e., `ST970900.xml` from station Strasbourg in September 1997) encoded in ESML is transformed via GRDDL to RDF/XML subject to the rules specified by the XSL file `ggpEsml2Rdf.xsl`. Because a number of GRDDL processor implementations have appeared in C, PHP, Python and XSLT [35], implementing the automated solution described here is a current focus. While Python-based GRDDL processors are anticipated to work best with the Python support provided with ESML [2], as of this writing, it's only the C-based Raptor parser that allows for input in XML [13] — all other processors support XHTML only.

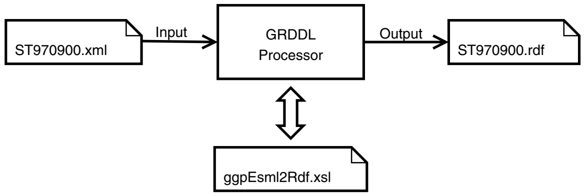

Figure 4. Simplified schematic for the RDF extraction from ESML via GRDDL. Note that URIs have been omitted in this figure.

Fig. 4 makes it explicitly clear that GRDDL is applied to per-station, monthly aggregates of GGP data. This has been the working assumption to this point. However, GRDDL could have been introduced earlier — i.e., to each of the `.GGP`, `.AUX` and `.LOG` files. This early introduction flow is illustrated schematically in Fig. 5. Note that the three, independent GRDDL conversions can proceed in parallel and are followed by a transformation involving XSLT.[10] In addition to merging the three, independent RDF representations of GGP data into a single monthly, per-station record, this transformation may also need to resolve redundancies (e.g., multiple instances of the same subject-predicate-object tuple) and/or inconsistencies (e.g., multiple subject-predicate-object tuples for a single subject) present in the RDF representation. Such representational refinements are quite likely to require processing that proceeds iteratively (P. Fox, personal communication). Finally, the degree of sophistication required here favors a transformative approach based on the richer-functionality XSLT rather than XInclude.

Although actual experience with GGP data will ultimately determine the most effective approach, early introduction of RDF should permit a better opportunity to:

- Capture and refine relationships in the GGP data — i.e., by making relationships clear at the most-granular level, prospects are improved for disam-

---

[9] As illustrated by Figs. 3 and 4, there exist different schematic representations for the GRDDL extraction of RDF from XML — and indeed the conversion of ASCII data into XML (Fig. 1). Review of the literature (e.g., [31, 32]) reveals further variation. The linear-systems-style schematic is used here owing to its familiarity in scientific contexts.

[10] An alternative approach has the following two steps: First, extract RDF from the `.GGP` and `.AUX` files as before. Second, incorporate data contained in the `.LOG` file via annotation. Annotation is a well-established practice [45, Chapter 4] involving RDF and the XML Pointer Language (XPointer, [22]) — essentially a URI-centric fragment identifier. This conversion flow is currently under investigation and the corresponding manuscript is in preparation.

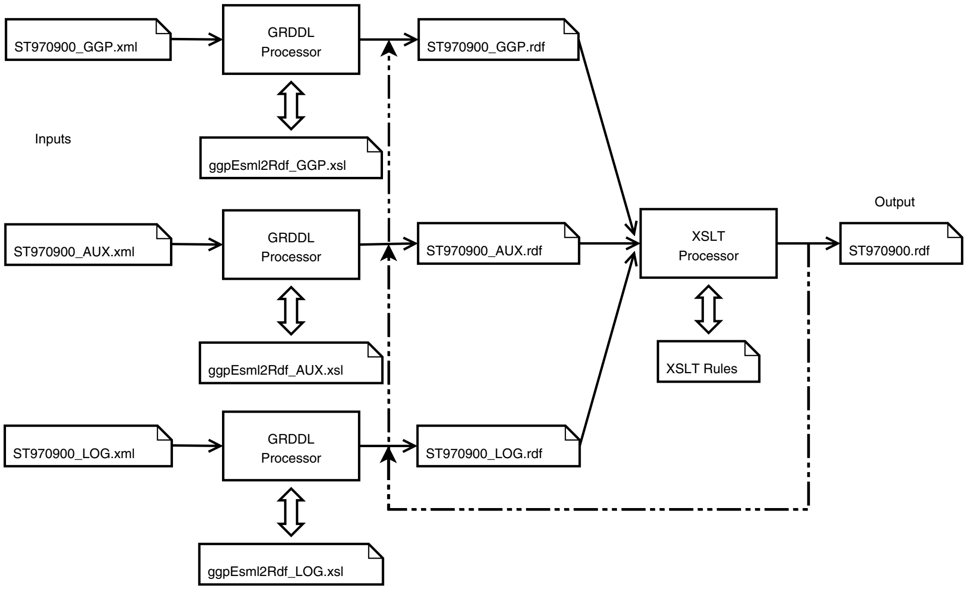

Figure 5. Early stage RDF extraction from ESML via GRDDL for the GGP. In a second stage, XSLT is used to combine the three separate and distinct RDF documents, and to perform additional iterative (dashed lines) processing (e.g., removal of redundancies and/or inconsistencies) as needed. As before, note that URIs have been omitted in this figure.

biguating meaning. This is a significant capability afforded by the flexibility of an automated RDF extrraction from ESML via GRDDL.

- Shape the specifics of the RDF representation — a capability whose value increases with the complexity of the XML representation, and/or when there exist multiple possibilities for extracting RDF from XML. For example in the case of the GGP, the ESML representation involves the routine use of multidimensional arrays. As such, it would desireable to shape the RDF representation in terms of RDF collections (e.g., [46, Chapter 4]) rather than explicit subject-predicate-object representations for each data point!

GRDDL shows significant promise in automating the extraction of RDF from XML from the bottom up. And although it has not yet been officially endorsed by the World Wide Web Consortium (W3C, [19]), it has been aligned with this organization's Semantic Web effort [20]. Thus it is an excellent time to get involved by offering feedback in the form of requirements to those developing the specifications in support of GRDDL [5, 35]. In addition to the standards front, GRDDL is being actively promoted [27] and existing implementations [5, 35] do exist. Given that this is the status of GRDDL today,[11] it is the current suggestion to compliment the bottom-up approach via GRDDL with a 'top-down' approach — a transformative process that starts at the highest level of abstraction and proceeds towards lower levels. The complimentary top-down approach is given attention in the following section.

4. Top-Down Introduction of RDF

In the previous section, VOC2OWL and GRDDL were identified as two examples of bottom-up approaches for enabling XML-to-RDF conversions. Although there are likely others, a more-exhaustive investigation will be required to identify them. In striking contrast, there are a number of identifiable, more-formal, top-down *ontological* approaches that are ultimately expressions in RDF. Because an "… ontology defines the common words and concepts (the meaning) used to describe and represent an area of knowledge …" [29], there has been considerable activity in particularizing and/or extending (e.g., [47]) generic ontologies (e.g., the Semantic Web for Earth and Environmental Technology, SWEET, [15]). These activities are significantly aided by 'ontology editors' such as Protégé [12] and SWOOP [17].[12]

As noted at the end of the previous section, GRDDL is in its earliest stages of development. Therefore, it is expected that a bottom-up extraction of RDF from XML via GRDDL

---

[11] In addition to the formal W3C sites, there are other sources for up-to-date developments with respect to GRDDL [18].

[12] In fairness, Protégé and SWOOP are much more than ontology editors: They are integrated development environments.

will be well complimented by top-down analysis and prototyping available from ontology tools such as Protégé and SWOOP. In other words, this *combined* bottom-up, top-down approach is anticipated to accelerate the rate of introduction of an RDF-based model for the GGP [41]. Use of these same ontology tools is also anticipated to assist in eventually harmonizing the integration of informal ontologies (e.g., that being derived here for the GGP) with SWEET-based and other formal ontologies [41].

5. The New Geoinformatics

In a previous paper, an XML-based data model was introduced for the GGP [38]. In the present paper, attention has focused on converting that representation into one based on RDF. Because the resulting RDF tuples can automatically populate a traditional database (e.g., via a script written in Python, [30]), it is very clear that this approach has the potential to address one of the identified scientific objectives [38] — i.e., the ability to identify GGP data on the basis of temporal and/or spatial specifications … to make it ‘findable’ [43, pg. 4].

Of course, this scientific imperative can be addressed by XQuery in the case of the XML-based model for the GGP ([38]). However, as the underpinning *lingua franca* of the emerging Semantic Web [29, 45], relationship-centric RDF will allow for more meaning-rich queries — e.g., via SPARQL [16]. The RDF introduction also establishes the basis for a much-richer data experience for the GGP [38, 41]. Phrased differently, the systematic introduction of context enhances semantic richness by transforming data into information, and information into knowledge. This well-established “stack of expressive power” [14, Slide 27] is clearly of value to the GGP in isolation [38]. However, even more compelling is the broader context within which GGP data now has currency and utility: “The ability to explain a train of reasoning may emerge as one of the most important capabilities a Semantic Web reasoning system can have” [45, pg. 131].

The recent, devasting 26 December 2004 Sumatra-Andaman earthquake [44] and resulting tsunami [36] provide one illustration of such a broader context. In such cases, there is no shortage of data leading up to, during or after the earthquake and tsunami. The data originates from multiple sources (e.g., scientific sensors, eye-witness accounts, etc.) of varying credibility, accuracy and quantitative specificity. GGP data is but one example of data available from a scientifically oriented sensor. To quote Geist et al. [33]:

> … this tsunami was the best documented in history — opening a unique opportunity to learn how to avoid such catastrophes in the future. From home videos of muddy water engulfing seaside hotels to satellite measurements of the waves propagating across the open ocean, the massive influx of information[13] has reshaped what scientists know in several ways.
>
> … this event revealed that subtle complexities of an earthquake exert a remarkably strong influence over a tsunami’s size and shape.[14]

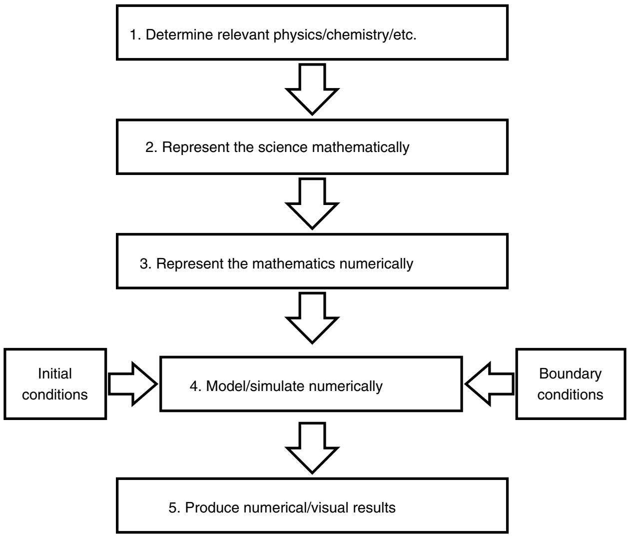

Figure 6. Schematics of the traditional (i.e., non-informatics) approach in the physical sciences (after [37]).

Despite this real-time data deluge, the ability to deliver early warnings of potential tsunami existence and impact plus tsunami physics in general, requires considerable improvement [33, 36]. Such high-impact opportunities, in tandem with inherent “subtle complexities”, amplify the need for a more-integrated approach.

The RDF-based approach described here permits GGP data to be readily assimilated into broader contexts — in the above tsunami-based example, to assist in the ability to produce actionable information or possibly knowledge. Because frameworks for the assimilation and analysis of data are appearing for use in contexts such as Homeland Security [25, 39, 49], there is much that can be applied to natural disasters such as tsunamis.

In essence, this approach posits the physical sciences as an ‘information science’. From this non-traditional vantage point, the physical sciences bear much in common with

---

[13] From the human perspective, Geist et al. make appropriate use of the term “information” here. However in the current context, in which processing by machines is implicit, the term “data” is more appropriate.

[14] An accompanying figure [33, pg. 63] effectively demonstrates this claim.

the bioinformatics area of the Life Sciences. The seminal discovery that DNA can be described and represented as an encoding scheme lead to a quantitative revolution, the likes of which the Life Sciences had not previously experienced. Even though the physical sciences have long enjoyed quantitative approaches based on mathematical representations of the relevant physics, chemistry, etc. (see Fig.6), challenges extant in areas such as natural disasters (e.g., tsunamis), call for a 'geoinformatic approach'. It is important to emphasize that this is really a call for a 'new geoinformatics', as the term 'geoinformatics' is already in use in the discipline to denote surverying and engineering, plus Geographic Information Systems (GIS). This perspective of the 'new geoinformatics' resonates fully with that espoused by leading-edge cyberinfrastructure initiatives such as GEON [3]. Sensor Data Fusion (SDF) also appears promising in its ability to provide a formal framework for the alliance of data originating from different sources [38, 50, 51]. Future investigations are needed to better understand these possibilities.

6. Conclusions

In a previous paper, reviewed here in §2, Lumb & Aldridge [38] introduced an ESML-based model for the GGP. Building on this XML-based foundation, and mindful of the need to incorporate metadata into the description and representation, the introduction of an RDF-based approach has been emphasized here (§3). Use of RDF's subject-predicate-object-tuple paradigm systematically allows relationships to be described and represented, and eventually result in an 'informal ontology' [41]. Key to this introduction from the bottom up is the recent arrival of GRDDL — a vehicle that extracts RDF from XML opposite rules. Because there exists some latitude in such extractions, a top-down approach was encouraged for the purpose of providing insight into the possibilities from a different perspective (§4). Integrated development environments like Protégé and SWOOP are of significant value in this top-down process. At this point, it is reasonable to expect that some effort will be required to reconcile the bottom-up and top-down approaches —- especially as these reconciliations involve formal ontologies. Ongoing investment in ontological approaches such as this allow GGP data to play in broader contexts where geoinformatics is the approach taken in addressing challenging scientific problems with the potential for significant human impact (§5). In hindsight, it is becoming increasingly clear that small-to-medium-scale projects like the GGP abound in interesting challenges to which solutions based on Grid Computing can be applied.

Acknowledgments

IL was introduced to GRDDL during Tim Berners-Lee's keynote [27] at the May 2005 Bio-IT World event in Boston. Was it *not* for this introduction, serendipitously engineered by Karen Sopuch (Platform Computing Inc.), this paper would not exist — and IL would *still* be pondering challenges extant in metadata. Peter Fox (The National Center for Atmospheric Science, NCAR, Boulder, Colorado, USA) shared the recursive nature of RDF extraction with KDA at the 2005 Fall Meeting of the American Geophysical Union (AGU) during a session convened on "Ontologies for Earth and Space Sciences". Finally, Volodya Savastiouk (Environment Canada, Downsview, Ontario, Canada) reminded us that XML's inherent ability to allow for validation was a useful way of ensuring the integrity of the data being represented. Each of these individuals is gratefully acknowledged for their interest and contribution to this investigation.

References

[1] Data Format Description Language Working Group. http://forge.gridforum.org/projects/dfdl-wg.
[2] Earth Sciences Markup Lanaguage (ESML). http://esml.-itsc.uah.edu.
[3] GEON — Cyberinfrastrucutre for the Geosciences. http:-//www.geongrid.org.
[4] GGP's Sumatra Data. http://www.eas.slu.edu/GGP/sum-atra.html.
[5] Gleaning Resource Descriptions from Dialects of Languages (GRDDL). http://www.w3.org/2004/01/rdxh/spec/.
[6] Global Geodynamics Project. http://www.eas.slu.edu/GGP.
[7] HTTP — Hypertext Transfer Protocol Overview. http:-//www.w3c.org/Protocols.
[8] Linked Environments for Atmospheric Discovery. http://-lead.ou.edu/.
[9] Marine Metadata Interoperability Project. http://www.-marinemetadata.org.
[10] Naming and Addressing: URIs, URLs, ... http://www.-w3c.org/Addressing.
[11] OWL Web Ontology Language Overview. http://www.w3c.-org/TR/owl-features.
[12] Protégé. http://protege.stanford.edu/.
[13] Raptor RDF Parser Toolkit. http://librdf.org/raptor/.
[14] Semantic Web Concepts. http://www.w3c.org/2005/Talks-/0517-boit-tbl/.
[15] Semantic Web for Earth and Environmental Technology (SWEET). http://sweet.jpl.nasa.gov/.
[16] SPARQL Query Language for RDF. http://www.w3c.org-/TR/rdf-sparql-query.
[17] SWOOP - Hypermedia-based OWL Ontology Browser and Editor. http://www.mindswap.org/2004/SWOOP/.
[18] Technorati GRDDL Tag. http://www.technorati.com/tag/-grddl.
[19] The World Wide Web Consortium. http://www.w3.org.

[20] World Wide Web Consortium's Semantic Web. http://www.w3.org/2001/sw.

[21] XML Inclusions (XInclude) Version 1.0. http://www.w3.org/TR/xinclude.

[22] XML Pointer Language (XPointer) Version 1.0. http://www.w3.org/TR/WD-xptr.

[23] XML Schema. http://www.w3.org/XML/Schema.

[24] XSL Transformations (XSLT) Version 1.0. http://www.w3.org/TR/xslt.

[25] N. R. Adam, V. Atluri, and V. P. Janeja. Agency interoperation for effective data mining in border control and homeland security applications. *The Fifth National Conference on Digital Government (dg.o 2004)*, Seattle, Washington, USA, May 2004.

[26] L. E. Bermudez and J. Graybeal. Speeding up ontology creation of scientific terms. *Eos Trans. AGU*, 86(52), 2005. Fall Meet. Suppl., Abstract IN41B-07.

[27] T. Berners-Lee. Semantic Web: Challenge and Promise. *Bio-IT World Conference + Expo 2005*, Boston, Massachusetts, USA, May 2005.

[28] T. Berners-Lee and M. Fischetti. *Weaving the Web: The Original Design and Ultimate Destiny of the World Wide Web by Its Inventor*. HarperCollins, 1999.

[29] M. C. Daconta, L. J. Obrst, and K. T. Smith. *The Semantic Web: A Guide to the Future of XML, Web Services, and Knowledge Management*. Wiley Publishing Inc., 2003.

[30] B. Ducharme. RDF: The Resource Description Framework. *Dr. Dobbs Journal*, 30(4):38–41, 2005.

[31] T. Erl. *Service-Oriented Architecture: A Field Guide to Integrating XML and Web Services*. Professional Technical Reference. Prentice Hall, 2004.

[32] T. Erl. *Service-Oriented Architecture: Concepts, Technology, and Design*. Professional Technical Reference. Prentice Hall, 2005.

[33] E. L. Geist, V. V. Titov, and C. E. Synolakis. Tsunami: Wave of Change. *Scientific American*, 294(1):56–63, January 2006.

[34] C. A. Goble, D. D. Roure, N. R. Shadbolt, and A. A. A. Fernandes. Enhancing Services and Applications with Knowledge and Semantics. In I. Foster and C. Kesselman, editors, *The Grid: Blueprint for a New Computing Infrastructure, Second Edition*, chapter 23, pages 431–458. Morgan Kauffman, 2004.

[35] D. Hazaël-Massieux. Bridging XHTML, XML and RDF with GRDDL. *XTech 2005: XML, the Web and Beyond*, Amsterdam, Netherlands, May 2005.

[36] C. Lomnitz and S. Nilsen-Hofseth. The Indian Ocean disaster: Tsunamic physics and early warning dilemmas. *Eos Trans. Am. Geophys. U.*, 86:65, 70, 2005.

[37] I. Lumb. HPC Grids. In A. Abbas, editor, *Grid Computing: A Practical Guide to Technology and Applications*, chapter 7, pages 119–133. Charles River Media, Inc., 2004.

[38] I. Lumb and K. D. Aldridge. Grid-enabling the Global Geodynamics Project: The introduction of an XML-based data model. In I. Kotsireas and D. Stacey, editors, *Proceedings of The 19th International Symposium on High Performance Computing Systems and Applications, HPCS 2005*, pages 216–222. The IEEE Computer Society, 2005.

[39] I. Lumb, B. McMillan, and M. Page. Task geometry for commodity Linux clusters and grids: A solution for topology-aware load balancing of coupled, asymmetric atmospheric models. In *Proceedings of the Eleventh ECMWF Workshop: Use of High Performance Computing in Meteorology*. World Scientific, 2005 (to appear).

[40] L. I. Lumb and K. D. Aldridge. Towards grid-enabling the Global Geodynamics Project. *Eos Trans. AGU*, 85(17), 2004. Joint Assembly Suppl., Abstract G34A-03.

[41] L. I. Lumb and K. D. Aldridge. Towards an ontology for the Global Geodynamics Project: Automated extraction of resource descriptions from an XML-based data model. *Eos Trans. AGU*, 86(52), 2005. Fall Meet. Suppl., Abstract IN43A-0328.

[42] S. P. Miller, D. Clark, J. Helly, D. Sutton, and T. Houghton. SIOExplorer: Advances across disciplinary and institutional boundaries. *Eos Trans. AGU*, 85(47), 2004. Fall Meet. Suppl., Abstract SF42A-08.

[43] P. Morville. *Ambient Findability*. O'Reilly & Associates, 2005.

[44] J. Park, K. Anderson, R. Aster, R. Butler, T. Lay, and D. Simpson. Global Seismographic Network records the great Sumatra-Andaman earthquake. *Eos Trans. Am. Geophys. U.*, 86:57, 60–61, 2005.

[45] T. B. Passin. *Explorer's Guide to the Semantic Web*. Manning Publications Co., 2004.

[46] S. Powers. *Practical RDF*. O'Reilly & Associates, 2003.

[47] R. Ramachandran and S. Movva. An ontology service for Linked Environments for Atmospheric Discovery (LEAD). *Eos Trans. AGU*, 86(52), 2005. Fall Meet. Suppl., Abstract IN43A-0330.

[48] D. D. Roure, N. R. Jennings, and N. R. Shadbolt. The Semantic Grid: A Future e-Science Infrastructure. In F. Berman, A. Hey, and G. Fox, editors, *Grid Computing: Making the Global Infrastructure a Reality*, chapter 17, pages 437–470. John Wiley & Sons, Ltd., 2003.

[49] SAP and R. University. Next-generation emergency response. *SAP SAPPHIRE 2005, SAP Innovation Pavilion*, Boston, Massachusetts, USA, May 2005.

[50] L. Wald. A European proposal for terms of reference in data fusion. *International Archives of Photogrammetry and Remote Sensing*, XXXII:651–654, 1998.

[51] L. Wald. Some terms of reference in data fusion. *IEEE Trans. Geosci. Rem. Sens.*, 37:1190–1193, 1999.