The 12th International Conference of
International Association for Computer Methods and Advances in Geomechanics (IACMAG)
1-6 October, 2008
Goa, India

International Data Exchange: The Future for Geo-engineering

D.G. Toll
School of Engineering, Durham University, Durham, United Kingdom

Keywords: data exchange; data representation; XML; case histories; case-based reasoning

ABSTRACT: The World Wide Web has revolutionised data access for all aspects of our daily lives. We are now able to find information quickly without leaving our desk. However, to allow full use of geo-engineering data from the web, as part of our professional activities, it is necessary for the data to be stored in a structured and standardised form. This will allow the World Wide Web to become an international repository for geotechnical information, available to the whole community. The paper discusses standardised XML schemes that are in development, presenting examples for borehole records, field test data and slopes. The paper also considers the use of a case-based reasoning system for slope design using a global database of slope case histories.

1 Introduction

The World Wide Web provides us with easy access to a huge amount of information. However, at present, the data we can access exists in many formats. To allow routine use of geo-engineering data from the web, as part of our professional activities, it is necessary for the data to be stored in a structured and standardised form. The way to achieve this is by adopting XML (eXtensible Markup Language) and developing internationally agreed data standards for geo-engineering.

XML is a simple and highly extensible way to represent data, which will allow data standards to continue to evolve to meet the needs of geo-engineering professionals. XML is a more generic form of mark-up language than HTML (Hyper-Text Markup Language), which has been the main language used on the World Wide Web. XML allows simple text files to be 'marked up' by including 'tags' that can be used to give meaning to the contents of a file; for instance data can be marked up using <borehole> ... </borehole> tags to indicate that all data between these tags relates to borehole information. These tags can be recognised by an XML compliant web browser. XML is being widely adopted by web developers for producing web-based materials (http://www.w3.org/XML/).

The concept of creating a geotechnical version of XML was first proposed by Mete Oner and the World Wide Web of Geotechnical Engineers (http://www.ejge.com/Gml/) in 1998. There are now a number of initiatives to develop representation schemes, both for geo-engineering and for geo-science data. The three international geo-engineering societies (International Society for Soil Mechanics and Geotechnical Engineering (ISSMGE), International Association for Engineering Geology and the Environment (IAEG) and International Society for Rock Mechanics (ISRM)) have a Joint Technical Committee, JTC2 (http://www.dur.ac.uk/geo-engineering/jtc2) to oversee the development of internationally agreed forms of representation of geo-engineering data. The data standards developed can be used to store such data on the World Wide Web and will ensure that geo-engineering data is stored in the same format anywhere on the web.

The major advantage of having data standards for web-based data is that it will make it possible to search all geotechnical data available on the web using structured search options (XQuery). For instance, it would be possible to locate XML files that contain projects within a particular geographical location or having particular soil/rock types or where a particular type of test has been performed. In this way the World Wide Web will become an international repository for geotechnical information, available to the whole community. This avoids the necessity to establish national or international geotechnical databases; each data owner can make their data directly available on their own web server.

A further advantage of developing data standards is to allow transfer of data between computer systems. XML uses very simple text files that can be easily accessed and read. Therefore, it can be used as a data exchange format between different organisations or an interchange format for linking different software packages. In this way XML could become the integrator between the different types of geotechnical tools (databases, knowledge-based systems, visualisation packages, conventional calculation software, numerical modelling packages etc.).

The paper discusses standardised XML schemes that are in development, with examples for borehole records, field test data and slopes. The paper also considers the use of a case-based reasoning system for slope design using a global database of slope case histories.

2 Data Exchange

The potential for a standard format for exchange of geotechnical data was recognised by Greenwood (1988). Threadgold and Hutchison (1992) identified the need to have a standard interchange package that was independent of particular software packages. This led to the development of the Association of Geotechnical and Geoenvironmental Specialists (AGS) data format in the UK in 1992 (the current version is available as AGS, 2004). A similar scheme was developed in the Netherlands (CUR, 2000). More recently, XML data formats have been proposed (Chandler et al, 2006; Weaver et al, 2008).

Weaver et al (2008) have identified the advantages of a standard data exchange format as being able to:

- exchange data between databases within an agency or organization
- receive data from consultants in a standard format
- exchange information with other agencies
- perform data quality checks
- exchange data between software packages and providers
- produce software products that are more standardized and more compatible with other products
- utilize and analyze data from various sources in an integrated geoenvironmental/ geotechnical data management system.

Such data exchange formats have been highly successful. Toll (1996) reported strong support for the AGS format from all involved, from data producer to data receiver/user, and also software developers. The format brought greater reliability and consistency to data transfer. However, such schemes were prone to problems when data sets were generated that did not fully conform to the standard. The use of XML has the advantage that it can be self validating; the data standard can have a "schema" definition that defines the required structure for the data and the naming of the tags. It can even contain validation rules that specify the required format of the data within the tags, or even limiting ranges for numerical values. A data file that does not conform to the standard will show errors when validated against the schema.

3 Data Representation

Rather than just developing a data exchange standard, Toll and Cubbitt (2003) recognised that XML would provide the means to store geotechnical data on the World Wide Web. They discussed how geotechnical entities (e.g. foundations, retaining walls and dams) could be represented in XML. Toll and Shields (2003) and Chandler et al (2006) also described XML schemes for representing ground investigation data. Hatipoglu (2003) outlined an XML schema for storing case histories of slopes (http://www.ins.itu.edu.tr/bulent/slopesml/) and this was extended by Toll (2007b). Styler et al (2007) described a scheme for deep foundations.

Although the data representing geotechnical entities is stored in a simple text-based XML file, the data can be presented in a number of ways in a web browser. The advantage of using XML to represent data is that the data (stored in an .xml file) is separated from the formatting information. Formatting is provided by the use of a Stylesheet (.xsl) file. This means that the data can be formatted in different ways for presentation without having to make changes to the data file. This separation between data and formatting instructions is a major advantage compared to HTML where formatting commands are embedded in the data file. Toll and Shields (2003) showed how XML data can be formatted using stylesheets or represented graphically using Java applets (Figure 1). The use of style sheets for representing slope case studies, boreholes and retaining walls can be seen at http://www.dur.ac.uk/geo-engineering/geotechml/ .

Toll (2007a) discusses some of the current initiatives underway to develop data standards for geo-engineering. The most significant development is the DIGGS project (Data Interchange for Geotechnical and Geo-environmental Specialists) (http://www.diggsml.org/). This is a collaboration between the Federal Highway Administration (FHWA), United States Environmental Protection Agency (US EPA), US Army Corps of Engineers, US Geological Survey (USGS), Eastern Federal Lands Highway Division (EFLHD) and a number of Departments of Transport in USA, funded through the Transportation Pooled Fund. The UK Highways Agency is also a collaborator. DIGGS brings together existing standards developed by AGS (www.ags.org.uk/agsml), Consortium of Organizations for Strong-Motion Observation Systems (COSMOS) (http://geoinfo.usc.edu/gvdc/) and the University of Florida, Department of Civil Engineering (http://fdot.ce.ufl.edu/).

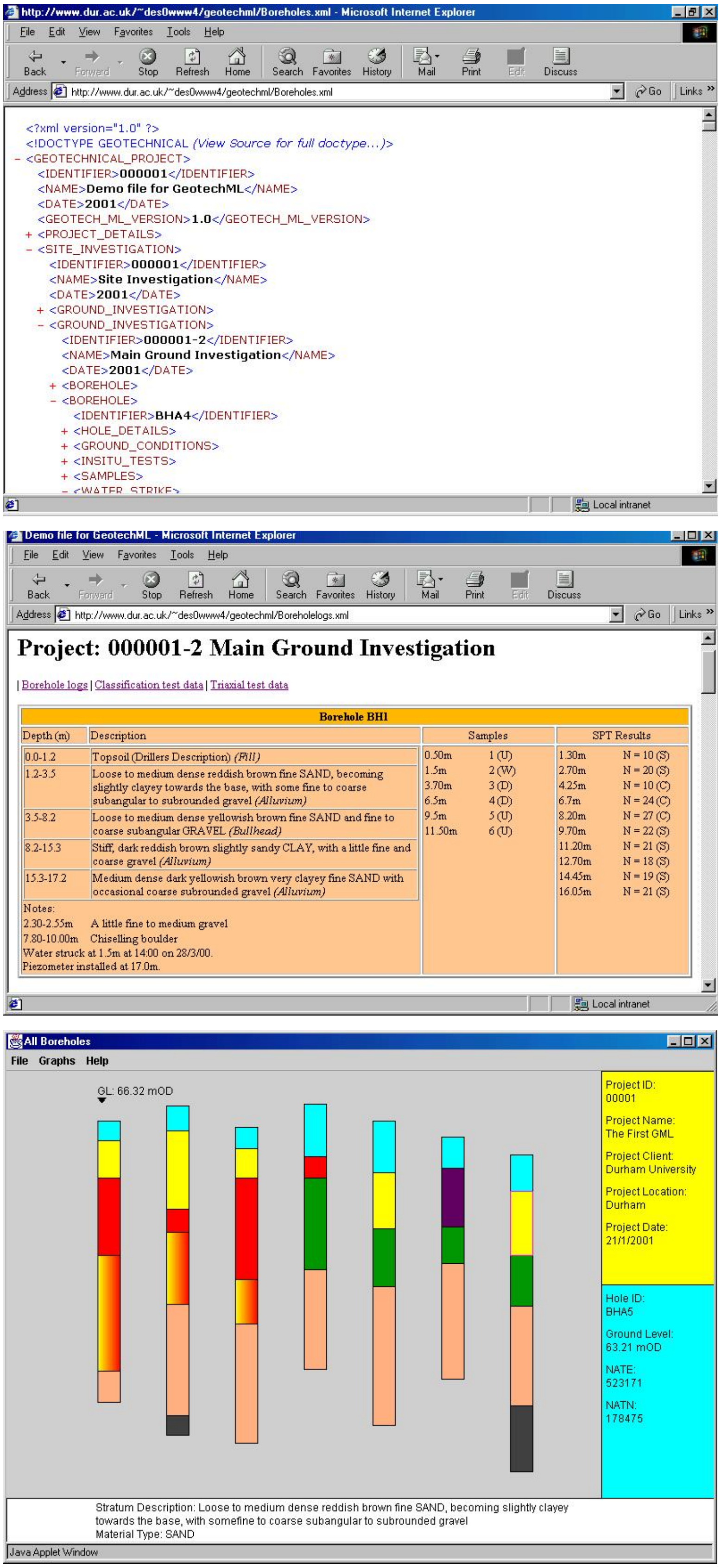


Figure 1. XML data displayed in different ways in a web browser (a) displayed as text (b) formatted as a borehole log using XSL (c) graphically represented as a cross-section using Java

In addition to Geo-Engineering schemes there have been parallel developments of XML schemes for geological sciences and mining. These include eEarth, XMML (eXploration and Mining Markup Langauge) and GeoSciML. eEarth (http://www.eearth.nl) was a European funded project that links the Geological Surveys of six European countries. The project aimed to increase availability, use and distribution of the digital subsurface data across Europe. The project is now completed and the website provides access to borehole information from several countries and in multiple languages (Figure 2). XMML (https://www.seegrid.csiro.au/twiki/bin/view/Xmml) is aimed at geoscience and exploration information. It is likely that this will be subsumed by GeoSciML (https://www.seegrid.csiro.au/twiki/bin/view/CGIModel/GeoSciML) which aims to represent geoscience information associated with geologic maps and observations, as well as being extensible in the long-term to other geoscience data.

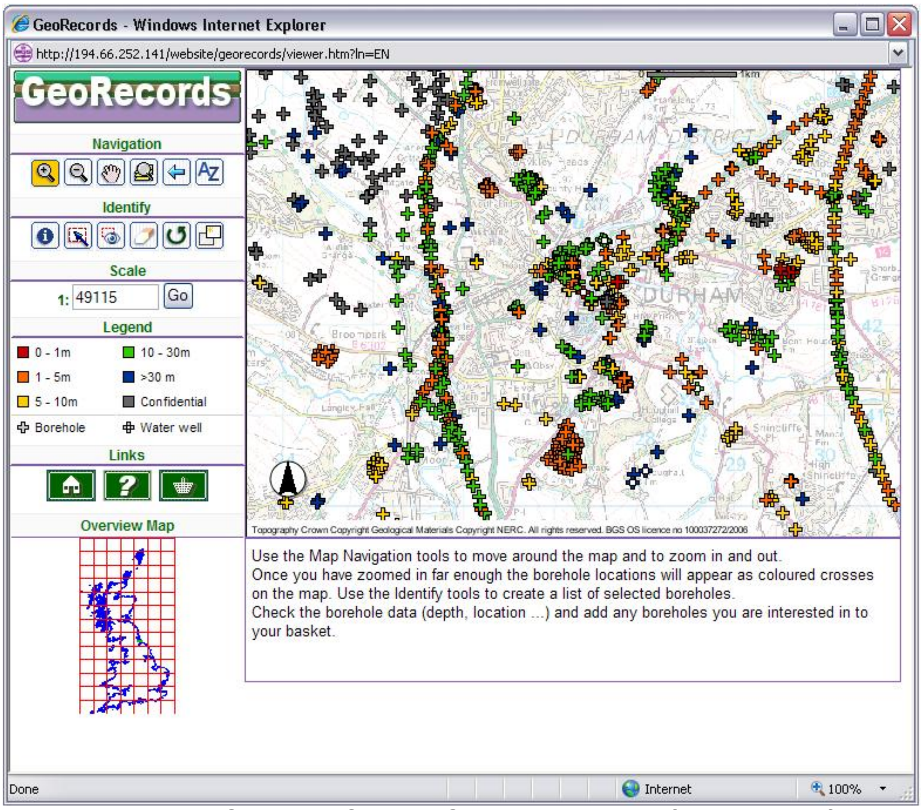


Figure 2. The British Geological Survey's GeoRecords system for borehole information available through eEarth (http://www.eearth.nl)

Toll (2007a) identifies the problem of a lack of consistency between geo-engineering schemes (such as DIGGS) and geo-science schemes (such as eEarth and GeoSciML). A common area is the representation of *boreholes* and *intervals* (represented as *hole/borings* and *layer/geology* respectively in the geo-engineering schemes). It is vital that these differences are eliminated to avoid a lack of compatibility between common data entities being used by both the geo-engineering and geo-science communities.

There have also been developments in generic representations schemes, using XML, that have wide application, such as Geography Mark-up Language (GML) and SensorML. GML (http://www.opengis.net/gml/) is a widely used and well-developed standard for representing geo-spatial data. It has constructs for representing coordinate schemes, locations and features that are geo-referenced. DIGGS has made extensive use of GML for representing spatial data. However, Toll (2007a) suggests that the location representation could use a simpler GML construct to aid compatibility with other schemes.

SensorML (2005) was initially developed to represent sensors for earth observation by remote sensing (http://vast.uah.edu/SensorML/). However, much of the work is generic enough to be applicable to other types of sensors. Toll (2008) shows that this generic format could be used to represent data from laboratory and field testing and gives an example of representing Cone Penetration Test data using SensorML constructs. It provides a very efficient data format for storing large amounts of numerical data. Figure 3 shows a comparison between AGSML and a proposed use of SensorML (Toll, 2008).

```
<Proj>
  <projId>GC2008</projId>
  <projEng>Durham University</projEng>
  <projDate>2007-02-18</projDate>
  <projAGS>3.1</projAGS>
  <holes>
    <Hole>
      <holeId>GEO-55</holeId>
      <fieldTesting>
        <Stcn>
          <Stcn_Dpth uom="#m">1.52</Stcn_Dpth>
          <Stcn_Res uom="#MPa">0.382</Stcn_Res>
          <Stcn_Fres uom="#MPa">0.0127</Stcn_Fres>
          <Stcn_Pwp1 uom="#kPa">-2.1</Stcn_Pwp1>
        </Stcn>
        <Stcn>
          <Stcn_Dpth>1.54</Stcn_Dpth>
          <Stcn_Res>0.382</Stcn_Res>
          <Stcn_Fres>0.0137</Stcn_Fres>
          <Stcn_Pwp1>4.5</Stcn_Pwp1>
        </Stcn>
        <Stcn>
          <Stcn_Dpth>1.56</Stcn_Dpth>
          <Stcn_Res>0.375</Stcn_Res>
          <Stcn_Fres>0.0151</Stcn_Fres>
          <Stcn_Pwp1>12.6</Stcn_Pwp1>
        </Stcn>
        <Stcn>
          <Stcn_Dpth>1.58</Stcn_Dpth>
          <Stcn_Res>0.375</Stcn_Res>
          <Stcn_Fres>0.0164</Stcn_Fres>
          <Stcn_Pwp1>19.1</Stcn_Pwp1>
        </Stcn>
        <Stcn>
          <Stcn_Dpth>1.60</Stcn_Dpth>
          <Stcn_Res>0.414</Stcn_Res>
          <Stcn_Fres>0.0169</Stcn_Fres>
          <Stcn_Pwp1>5.1</Stcn_Pwp1>
        </Stcn>
      </fieldTesting>
    </Hole>
  </holes>
</Proj>                                      (a)
```

```
<swe:Data>
  <swe:definition>
    <swe:DataDefinition>
      <swe:dataComponents>
        <swe:DataGroup>
          <swe:component name="conePenetration">
            <swe:Quantity definition="Penetration Length"
                          uom="#m"/>
          </swe:component>
          <swe:component name="coneResistance">
            <swe:Quantity definition="Cone Tip Resistance"
                          uom="#MPa"/>
          </swe:component>
          <swe:component name="coneFriction">
            <swe:Quantity definition="Friction Sleeve Resistance"
                          uom="#MPa"/>
          </swe:component>
          <swe:component name="conePwp1">
            <swe:Quantity definition="Pore Water Pressure 1"
                          uom="#kPa"/>
          </swe:component>
        </swe:DataGroup>
      </swe:dataComponents>
      <swe:encoding>
        <swe:AsciiBlock decimalSeparator="."  tokenSeparator=","
              tupleSeparator=" "/>
      </swe:encoding>
    </swe:DataDefinition>
  </swe:definition>
  <swe:value>
    1.52,0.382,0.0127,-2.1
    1.54,0.382,0.0137,4.5
    1.56,0.375,0.0151,12.6
    1.58,0.375,0.0164,19.1
    1.60,0.414,0.0169,5.1
  <swe:value>
</swe:Data>                                       (b)
```


Figure 3. Comparison of (a) AGSML and (b) SensorML for representing Cone Penetration Data

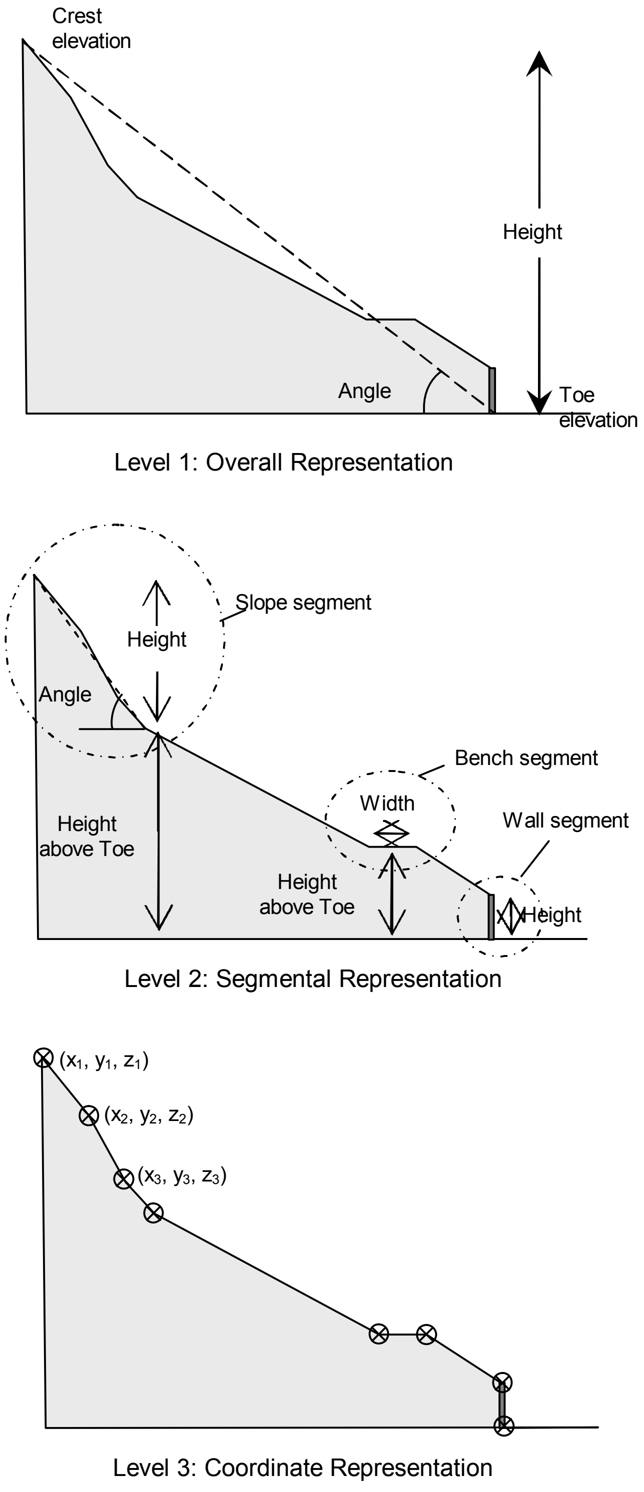
4 Representing Slopes and Landslides

Slope data is used by a range of professionals: geotechnical engineers, geomorphologists, geologists and planners. They may each need to represent information in different ways. Even if we consider only topographic information, different levels of representation are likely to be needed. Those dealing with hazard assessment will typically store data on the slope height, slope angle and aspect (as well as lithology and land use). Geomorphologists may want to divide a slope up into segments having different land forms. Geotechnical engineers will usually produce quantitative cross-sections showing detailed topography.

It is therefore essential that a representation scheme is capable of operating at any of these different levels of representation. Toll (2007b) outlined a three level representation of topography that can be used to satisfy the varying requirements (Figure 4). At Level 1 the slope is defined simply by an overall slope angle and overall slope height (or Crest and Toe elevations). At Level 2 the slope is divided into segments (each referenced by height above the toe) which can be described as Slope Segments (defined by angle/height), Bench Segments (primarily defined by bench width) or Wall Segments (primarily defined by wall height). Of course, each segment can have other properties attached, such as surface cover or geomorphological descriptions. At Level 3 the topography is defined by coordinates allowing a very detailed topographic representation (in 2D or 3D).

Fyson and Toll (2007) demonstrated a case-based reasoning system (SlopeSafe) for preliminary design of slopes. SlopeSafe draws on a case-base of records of successful and failed slopes to give an indication of the likely success of a proposed slope by matching its geometry and ground conditions to the slopes held in the

The 12th International Conference of
International Association for Computer Methods and Advances in Geomechanics (IACMAG)
1-6 October, 2008
Goa, India




```
- <Slope>
  Level 1: Overall Data
  - <Toe>
      <Location />
      <Elevation />
    </Toe>
  - <Crest>
      <Location />
      <Elevation />
    </Crest>
  - <SlopeGeometry>
      <SlopeAngle />
      <Height />
      <ProfileShape>Concave/Linear/Convex</ProfileShape>
      <PlanShape>Concave/Linear/Convex</PlanShape>
    </SlopeGeometry>
  - <Aspect>
      <Direction>N/NE/E/SE/S/SW/W/NW</Direction>
      <Bearing />
    </Aspect>
  - <Upslope>
      <SlopeAngle />
      <Height />
      <Feature>Road/River/Building etc</Feature>
      <FeatureDistance />
    </Upslope>
  - <Downslope>
      <SlopeAngle />
      <Height />
      <Feature>Road/River/Building etc</Feature>
      <FeatureDistance />
    </Downslope>
  Level 2: Segmental Data
  - <SlopeSegments>
    - <SlopeSegment>
        <SegmentID />
        <HeightAboveToe />
        <SlopeGeometry />
        <SurfaceCover />
      </SlopeSegment>
    - <BenchSegment>
        <SegmentID />
        <HeightAboveToe />
        <BenchWidth />
      </BenchSegment>
    - <WallSegment>
        <SegmentID />
        <HeightAboveToe />
        <WallType />
        <WallAngle />
      </WallSegment>
    </SlopeSegments>
  Level 3: Coordinate Data
  - <SlopeCoordinates>
    - <GroundSurface>
        <Geometry />
      </GroundSurface>
    - <PhreaticSurfaces>
        <SurfaceID />
        <Geometry />
      </PhreaticSurfaces>
    - <GroundLayers>
      - <Layer>
          <LayerID />
          <Geometry />
          <GroundConditions />
        </Layer>
      </GroundLayers>
    - <RuptureSurfaces>
        <SurfaceID />
        <Geometry />
      </RuptureSurfaces>
    </SlopeCoordinates>
  </Slope>
```


Figure 4. Three levels of representation for the same slope

case-base (Figure 5). XML has been used to create the case-base. The system, as originally designed, used a local XML file containing almost 3000 case records. However, slope case histories could be made available on the web, using a standard slope representation scheme extending that reported by Toll (2007b) (http://www.dur.ac.uk/geo-engineering/geotechml/Slopes/CaseHistoryInventory.xml). In this way, it would be possible for a case-based design system to access an international case-base of slope case records, by performing an XQuery search. Such a design system would automatically draw on new experience every time a new case history was added to the web, anywhere in the world.

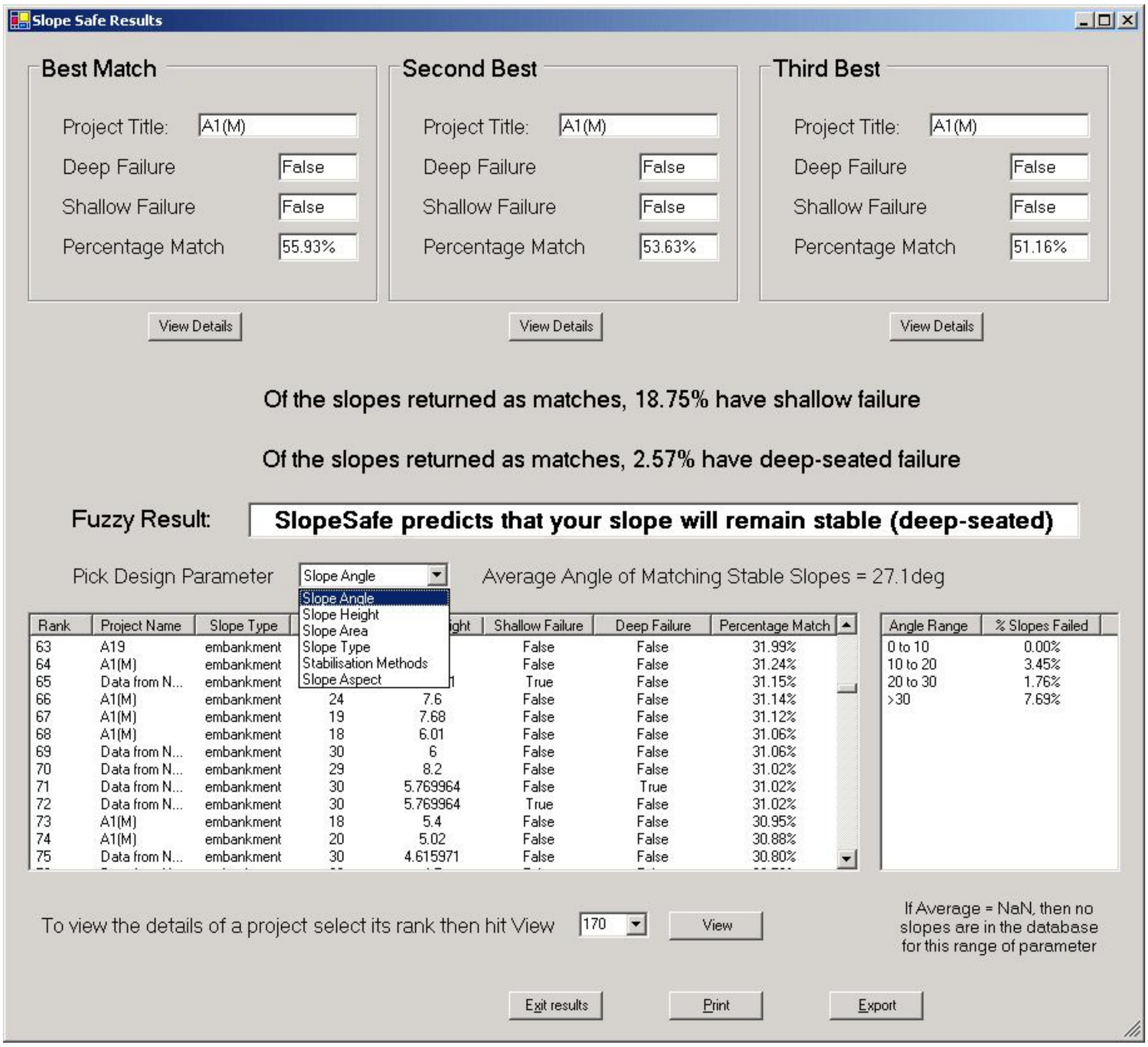


Figure 5. Results screen from the SlopeSafe case-based reasoning system (Fyson and Toll, 2007)

5 Conclusions

It is necessary for geo-engineering data to be stored on the World Wide Web in a structured and standardised form if it is to provide a useful repository of data. The way to achieve this is by adopting XML (eXtensible Markup Language) and developing internationally agreed data standards for geo-engineering. There are now a number of representation schemes in development for geo-engineering and geo-sciences. At present, there is lack of consistency between geo-engineering and geo-science schemes and it is vital that these differences are eliminated to avoid incompatibility between common data entities that are used by both the geo-engineering and geo-science communities.

The advantage of using XML to represent data on the World Wide Web is that the data is separated from the formatting information. This separation between data and formatting instructions is a major advantage compared to HTML where formatting commands are embedded in the data file. The XML data can be viewed in a variety of ways within a web browser, either formatted using stylesheets or represented graphically using Java applets.

The use of a standardised XML data representation scheme will make the World Wide Web into an international repository for geotechnical information, available to the whole community. Such developments will make possible the use of case-based reasoning systems using global databases of case histories. Such a design system would automatically draw on new experience every time a new case history was added to the web, anywhere in the world.

6 References

AGS. 2004. Electronic Transfer of Geotechnical and Geoenvironmental Data, Edition 3.1, Association of Geotechnical and Geoenvironmental Specialists, Beckenham, Kent (with addendum May 2005) (available from http://www.ags.org.uk).

Chandler, R.J., Quinn, P.M., Beaumont, A.J., Evans, D.J. and Toll, D.G. 2006. Combining the Power of AGS and XML: AGSML the Data Format for the Future", *Proc. GeoCongress 2006: Geotechnical Engineering in the Information Technology Age*, Atlanta, USA, ASCE, Reston, pp. 112-117.

CUR. 2000. Description of the GEF Language, *Civieltechnisch Centrum Uitvoering Research en Regelgeving*, Gouda (available from http://www.geonet.nl/software/gef).

Fyson, M. and Toll, D.G. 2007. Case-Based System for Slope Design, *Computers and Geotechnics*, http://dx.doi.org/10.1016/j.compgeo.2007.06.007.

Greenwood, J.R. 1988. Developments in computerised ground investigation data, *Ground Engineering*, Sept. 1988, pp. 36-41.

Hatipoglu, B. 2003. The Development of an Integrated and Intelligent Design Environment for the Investigation of Slope Stability Problems, PhD Thesis, Istanbul Technical University.

GML. 2004. Geographic information – Geography Markup Language (GML) Version 3.1.1, *Report ISO/TC-211/WG-4/PT-19136*, International Organisation for Standardisation (http://www.iso.org/).

Styler, M., Hoit, M., and McVay, M. 2007 Deep Foundation Data Capabilities of the Data Interchange for Geotechnical and Geoenvironmental Specialists (DIGGS) Mark-up Language, *Electronic Journal of Geotechnical Engineering*, http://www.ejge.com/2007/Ppr0772/Ppr0772.pdf .

SensorML. 2005. Sensor Model Language (SensorML) Implementation Specification, *Report OGC 05-086*, Open Geospatial Consortium Inc. (http://www.opengeospatial.org/).

Threadgold, L. and Hutchison, R.J. 1992. The Electronic Transfer of Geotechnical Data from Ground Investigations, *Colloque International Géotechnique et Informatique*, Paris, Presse de l'Ecole National des Ponts et Chaussées, pp. 749-756.

Toll, D.G. 1996. Data Management: Discussion Reviewer's Report for Session II, in *Advances in Site Investigation Practice* (ed. C. Craig), Thomas Telford, London, pp. 250-256.

Toll, D.G. 2007a. Geo-Engineering Data: Representation and Standardisation, *Electronic Journal of Geotechnical Engineering*, http://www.ejge.com/2007/Ppr0699/Ppr0699.htm .

Toll, D.G. 2007b. Representing Slopes in XML, *Proc. Specialized Session S02 of 11th Congress of the International Society for Rock Mechanics*, Lisbon, pp. 13-17 (on CD) (http://www.dur.ac.uk/geo-engineering/jtc2/ISRM2007/ISRM_Specialised_Session_S02.pdf).

Toll, D.G. 2008. Representing Geo-Engineering Data from Instruments and Transducers, *Proc. GeoCongress 2008*, New Orleans, March 2008.

Toll, D.G. and Cubitt, A.C. 2003. Representing Geotechnical Entities on the World Wide Web, *Advances in Engineering Software*, Vol. 34, 11-12, pp 729-736.

Toll, D.G. and Shields, R. 2003. A Web-based Data Format for Ground Investigation Data, *Electronic Journal of Geotechnical Engineering*, http://www.ejge.com/2003/Ppr0321/Ppr0321.htm .

Weaver, S.D., Lefchik, T., Hoit, M. and Beach, K. 2008. Geoenvironmental and Geotechnical Data Exchange: Setting the Standard, *Proc. GeoCongress 2008*, New Orleans, March 2008.