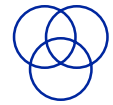
Esprit 25110

XML-based Interchange Format for STEP Data

Monsell EDM



Engineering Data Management

Robin La Fontaine,
STEPWISE Project Manager
Monsell EDM Ltd.
robin@monsell.co.uk
http://www.stepwise.org

Contents

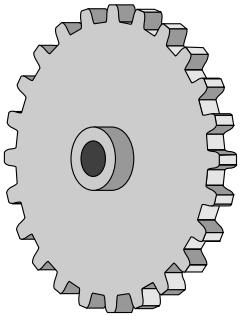
- STEP in brief
- How can XML be used for product data?
- What does STEP require from XML?
- Some technical stuff
- STEP and XML: A marriage made in heaven?

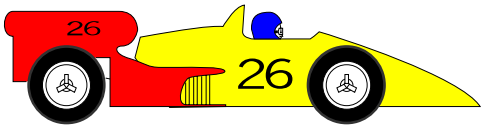
Esprit 25110

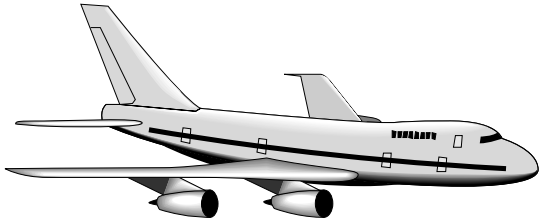
The STEP world: a brief introduction

Monsell EDM
Engineering Data Management

- Product data: a single component to a complete process plant, taking in printed circuit boards en route





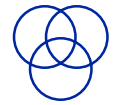


- Formal data (information) models using EXPRESS language
- Methodology for development and testing of the models
- Exchange format using ‘Part 21’

Esprit 25110

What is EXPRESS?

Monsell EDM



Engineering Data Management

- EXPRESS is a powerful information modelling language
- Covers data types
  - string, integer, boolean … complex structures
- Covers structural aspects
  - references and relationships
- Covers complex constraints
  - assertions about what is ‘good’ data

EXPRESS: an information modelling language (1)

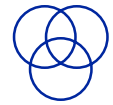
- Covers data types
  - string, integer, boolean … complex structures

```
TYPE
  temperature  = REAL;
END_TYPE;

TYPE
  plant_name   = STRING;
END_TYPE;

TYPE
  flower_colour  =   ENUMERATION OF (red, yellow, white);
END_TYPE;
```

Esprit 25110

EXPRESS: an information modelling language (2)

Monsell EDM



Engineering Data Management

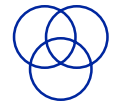
- Covers structural aspects
  - References and relationships

```
ENTITY garden;
  has_greenhouse                : greenhouse;
  climatic_temperature_range : temperature_range;
  has_beds                          : SET [5 : 5] OF bed;
END_ENTITY;

ENTITY greenhouse;
  enforced_temperature_range  : temperature_range;
  holds_plants                      : SET [1 : ?] OF greenhouse_plant;
INVERSE
  the_garden                 : garden FOR has_greenhouse;
END_ENTITY;
```

Esprit 25110

EXPRESS: an information modelling language (3)

Monsell EDM



Engineering Data Management

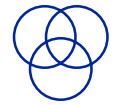
- Covers complex constraints
  - Assertions about what is 'good' data

```
TYPE
  ph    = REAL;
WHERE
  the_ph_is_between_0_and_14 : {0 <= SELF <= 14};
END_TYPE;

ENTITY outdoors_plant  SUBTYPE OF (plant);
  survival_ph_range          : ph_range;
WHERE
(*The ph range of the outdoors plant must include the ph value of  the bed *)
r1: QUERY (b <* the_beds | value_is_within_range (b.acidity, survival_ph_range))
              = the_beds;
END_ENTITY;
```

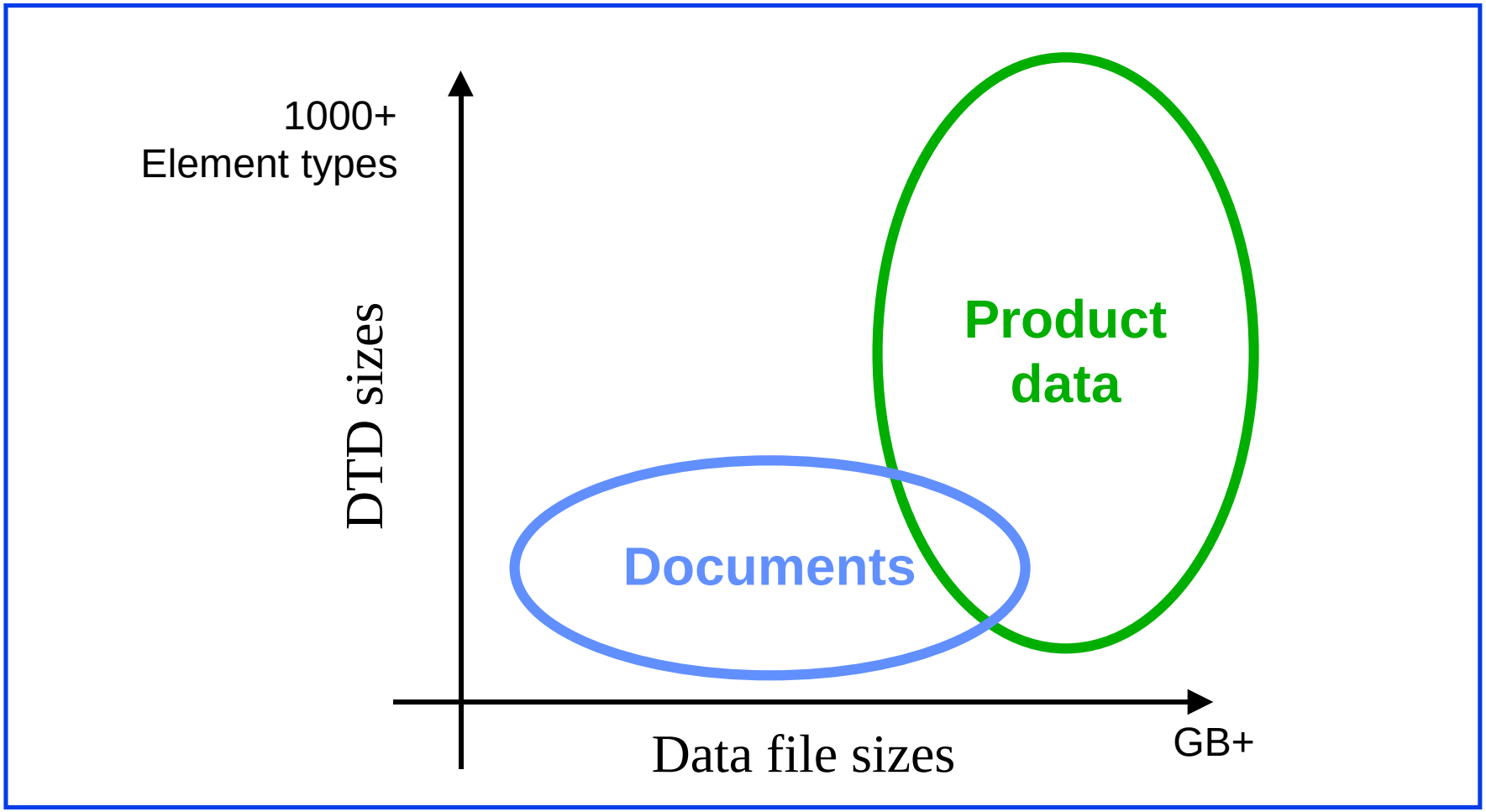
Esprit 25110

Use of XML within STEP: scale of the problem

Monsell EDM



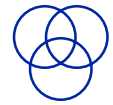
Engineering Data Management



Esprit 25110

Use of XML within STEP: application areas

Monsell EDM



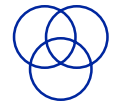
Engineering Data Management

- EXPRESS language
  - To allow a schema to be sent with the data
  - Allows easier access to the model using standard XML parsers
  - Gives robust mechanism for extensions and additional properties
- Exchange format for data
  - Better structures possible than with Part 21
  - More human-readable
  - Early late binding formats possible
  - Easier to generate and read for non-EXPRESS systems

Esprit 25110

Using XML as a serialisation format

Monsell EDM



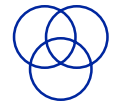
Engineering Data Management

- It does not solve all your problems!
  - "Your data is in XML, I can read XML, therefore I can read your data."
  - Not so!
  - You still need to understand the objects you are exchanging, and write software that processes them
- But the available technology and software makes it easier
  - Parsers (SAX), filters (XSL, DSSSL), standard API (DOM)

Esprit 25110

Challenges of an Interchange format

Monsell EDM



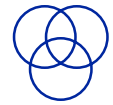
Engineering Data Management

- Requirements are more challenging than for a serialisation format
  - Human-readable and computer-readable of course
  - Strict syntax to ensure data is as correct as possible (ideal but unlikely: syntactically correct file is guaranteed semantically correct)
  - Ability to upgrade in upward-compatible manner as model is changed (so old data can be read into new readers)
  - Program-friendly: easy to generate/read the format from different programs (otherwise it is not much use as an exchange format!)
    - Avoid numeric pointers and object-Ids
    - Sensible referencing mechanism
    - Appropriate use of hierarchy

Esprit 25110

STEP's existing format, Part 21

Monsell EDM



Engineering Data Management

- In reality a serialisation format rather than an interchange format

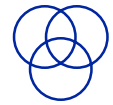
```
#29=COORDINATED_UNIVERSAL_TIME_OFFSET(8,$,.BEHIND.);
#45=APPROVAL(#46,'Approved as initial STEP test case part');
#46=APPROVAL_STATUS('approved');
#47=APPROVAL_DATE_TIME(#48,#45);
#48=DATE_AND_TIME(#50,#51);
#50=CALENDAR_DATE(1993,17,7);
#51=LOCAL_TIME(13,29,52.0,#29);
```

  - You need full knowledge of the EXPRESS model to understand the format (if the model changes, the format does also in non-upward compatible manner: think of all that test data that needs revision!)
  - Extensive use of numeric object-Ids, which are not persistent
  - No hierarchy in data file, just a big list of objects

Esprit 25110

So, what do we want? Requirements (1) …

Monsell EDM



Engineering Data Management

- Provide both early and late bound solutions (discussed later)
- Enable subsets of models to be transmitted
  - An agreed selection of the entities which the model defines (known as a ‘conformance class’ in STEP terminology)
- Allow transmission of entity sets from different models in one file, or one set in multiple files
- Allow choice between short files and more readable files
- Enable deltas (changes) to be represented and sent

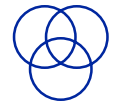
Requirements (2) …

- Enable multiple (human) languages
- Enable upward-compatible changes as model is developed
- Allow incorrect or incomplete data to be sent
- Enable as many semantic constraints as possible to be captured by syntax, e.g. referential integrity

Esprit 25110

Early and late binding solutions

Monsell EDM



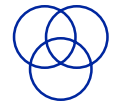
Engineering Data Management

- Late binding is a single format for all models
  - Easy apply generic processing to models
  - Difficult to apply model-specific processing
  - Good for data warehouse type applications
- Early binding is a specific format for each model
  - Bigger DTD than late bound (thousands of element types)
  - Generic processing of models difficult
  - Model-specific processing easier
  - Good for data exchange and display
- Can we get the best of both worlds?
  - SGML Architectures

Esprit 25110

Use of SGML Architectures

Monsell EDM



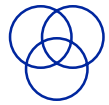
Engineering Data Management

- Allow us to relate an early-bound format to a late-bound view of the data
- Each early-bound, specific element types in the DTD will have attributes that give its representation as generic late-bound elements
  - There will be limitations as to the extent of the mapping that is possible
- Would allow more than one early-bound representation for a model
  - So you can use a format suited to your application which others can still read

Esprit 25110

DTD development principles

Monsell EDM



Engineering Data Management

- Use meaningful keywords (and gzip to compress!)
- Use names from EXPRESS model (or syntax) in DTD
- Prefer element content over attributes
- Wrap repeating sequences in containing element
- Use prefix notation for expressions

<lessthan><arg>a</arg><arg>b</arg></lessthan>

as opposed to

<func><arg>a</arg><lessthan/><arg>b</arg></func>

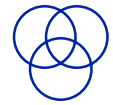
Other initiatives

- MathML
  - XML format for mathematical expressions
  - EXPRESS uses expressions
  - Perhaps we can use MathML in the XML format for EXPRESS
- XMLSchema
  - Scope much more limited than EXPRESS
  - Benefits of EXPRESS features has been submitted to XML Schema
- XMI
  - UML modelling language format for model and data
  - It does overlap
  - Needs more work to determine relationship

Esprit 25110

Relative strengths of STEP and XML

Monsell EDM



Engineering Data Management

- EXPRESS is a powerful information modelling language
  - Covers data types (string, integer, boolean … complex structures)
  - Covers structural aspects (references and relationships)
  - Covers complex constraints (assertions about what is ‘good’ data)
- XML is a file format
  - Widely available and understood
  - DTD allows some definition of structure
  - Easy to read and parse
  - Basic structure (list) is powerful and flexible
- It is a good match!

Esprit 25110

Finally...

Monsell EDM
Engineering Data Management