

The role of Standardisation and Interoperability in the IT sector with a focus on the reliance on XML.

Work in progress on Project for Computer
Mediated Communication (CE_6014)

Dave Healy.

XML

- **Xml is currently the standard for markup languages. However, it has a number of disadvantages.**
- **It is often seen as being quite verbose which leads to slow performance and data delivery.**
- **Wasteful in how much space it needs to use for the amount of true data that it is sending.**

The main advantages of XML

- While XML has big problems as a data-interchange format, these disadvantages are compensated for by the benefits of interoperability and support.
- The main reasons these advantages exist is that XML has been the standard for markup languages. If a more efficient markup language has been established as the standard, that different language would have these advantages also.
- The fact that XML has these advantages acts as a sort of barrier to entry for other markup languages.

The main alternatives to XML

- JSON – JavaScript Object Notation
- YMAL – YMAL Ain't Markup Language
- GPB – Google Protocol Buffers

JSON - JavaScript Object Notation

- **JSON** is a lightweight data-interchange format.
- It is easy for humans to read and write. It is easy for machines to parse and generate.
- Much better suited to data interchange than XML while holding most of the advantages of XML.
- JSON has a much smaller grammar and maps more directly onto the data structures used in modern programming languages than XML.

YAML – YAML Ain't Markup Language

- **YAML** is a straightforward machine parsable data serialization format designed for human readability and interaction with scripting languages such as Perl and Python.
- Representing simple, hierarchical data tends to be more gracefully done in **YAML** rather than **XML**, but actual marked-up text is awkward to represent.
- Intended to be a “cleaner version” of **JSON**.

GPB – Google Protocol Buffers

- Very dense and efficient data. However, this comes at the expense of not being easily read by humans
- Hard to robustly decode without knowing the schema (data format is internally ambiguous, and needs schema to clarify)
- Very fast processing due to the efficient layout.

Further work to be done

- Now that I have put in some research into the background of markup languages, I intend to examine the advantages and disadvantages of each and determine what exactly causes the problems of XML as a language.
- A recurring attribute is the trade off between a language being machine readable and Human readable. One reason XML is so inefficient is that it puts a bearing on being human readable which results in a lot of code that is useless when it comes to machine to machine interactions. As a side note, the fact that it is so easily read by humans is probably why it was preferred as the standard in the first place.
- GPB leans towards the machine readable side while JSON does a good job in balancing the two. I will be putting an emphasis on examining JSON to find out exactly how it balances the two and whether or not there are any disadvantages to this.



Thank you for listening