

University of St Andrews  
School of Computer Science

**Distinguished Lecture Series 2001/02**

*Wednesday 12<sup>th</sup> December 2001  
Lecture Theatre C, Mathematical Institute,  
North Haugh St Andrews*

**XML - a data standard for well-behaved  
programmers?**

by

**Professor Richard Connor**  
**University of Strathclyde**

Richard Connor is a Professor of Computer Science at the University of Strathclyde. The interests of his research group include querying semi-structured data resources, autonomous information provision, and highly distributed data stores. His work on distributed information is currently funded by EPSRC, BBSRC, SHEFC, and a global supplier of financial information systems.

School of Computer Science, University of St Andrews, North Haugh, St Andrews, Fife KY16 9SS.  
(Tel No. 01334 463253. Details on our web page: <http://www.dcs.st-and.ac.uk/CompSci/Teaching/dislec.html#dislec.2>)

# Abstract

"XML" is a term generally used to refer to many rapidly emerging standards and technologies. While XML 1.0 is stable, much of the associated set of standards and software is in a very early stage of development. The talks will examine emerging XML-related technology from the perspective of data and applications. While the WWW is the largest high-quality data resource ever to have been available to mankind, it lacks the inherent discipline thought to be necessary for a coherent programming framework. This lack is not necessarily a case of poor design, but rather reflects the inherent autonomy necessary for such a large collection to exist.

One lesson that the emergence of the Web has led many computer scientists to take is that groups of humans can be well-behaved to create major systems that work by convention, rather than by enforcement. XML captures a data model that can be viewed as a database standard, so long as generators and programmers stick to the conventions. The series of talks will examine this viewpoint.

The talks will be accessible to anyone who understands the basics of computer programming and simple type systems; no background knowledge of XML is necessary.

# Programme

- 10.00 – 11.00**    **When is a Document not a Document? When it's XML.**  
The first talk gives a background to the XML standard set, examining the re-emergence of the semi-structured data model and how that fits with the historical coincidence of XML. The value of XML as a data standard, despite its humble origins as a document standard, is highlighted. The value of such a data standard in a context subject to autonomy and evolution is explained.
- 11.00 – 11.30**    *Coffee*
- 11.30 – 12.30**    **When is a Language not a Language? When it's well-typed.**  
A background to the differing emerging paradigms is given, using a classification based upon program safety. XML does not have a strong type model; one of the most interesting questions is how programmers can write applications that will not fail in unexpected ways at run-time. The class of failure examined in detail is that which could be detected at compile-time using more standard database technology.
- 14.15 – 15.15**    **When is a SNAQue not a Snake? When it's a type projection.**  
Based on the classification given previously, we consider how standard data types may be projected onto XML collections, and show how for some purposes this approach is more promising than others in terms of program safety. The approach outlined is that of the Strathclyde Novel Architecture for Querying eXtensible Markup Language.