1. Introduction

Web Data and the two cultures:

- WEB provides a simple and universal standard for the exchange of information.

Information is decomposed into named units (URLs corresponding to a file) and transmitted.

HTML is the language used to structure the test for visual presentation (able to describe intra-document structure - layout and inter-document structure - hyperlinks).

HTTP is the protocol used by Web servers and Web Clients to exchange information (typically HTML documents).
of this data, but has access only to the HTML pages.

A second organization wants to obtain financial analyses
correlating the results into HTML.

are generated on demand by executing queries in SQL and
are retrieved from a relational database as the source. Web pages
Web using a relational database as the source. Web pages
An organization publishes financial information on the

Consider the following scenario:

- The need for a BRIDGE between the two cultures.

- An ad-hoc manner.

Query Languages are used to access the information in

data using strict structure definitions (schemas).

DATABASES provide mechanisms to represent and manipulate
SQL query.

This could have been accomplished easily with an
quantity such as an average of a single column.
Repeated requests for HTML pages to compute some
a large part of the underlying database through
(2) The software may have to download

the parsing software could break the parsing software
(1) A minor change in the HTML formatting

Problems with this approach:

for the data analysis software.

parses HTML data and converts into a suitable format
One solution: Second organization writes software that
MIDDLE TIER responsible for transforming/interfacing/adapting value
several CLIENTS/CONSUMERS of the data
several SERVERS/DATA SOURCES

traditional 2-tier (client/server) to multi-tier

Shift in paradigm:

two cultures.
capability is a step towards the convergence of the a data model for semi-structured data (XML) with querying of the display format; easily transmitted and exchanged.
XML provides ability to structure data independently.

two views.

XML is the first step towards the convergence of the