	<b>Project Note</b>	Version 01	Status: Issued	Date 23-Jun-05
	Project: JCLD Pilot Project in Anticipation of E-Journals		Topic: Project Report - June	


## Project Report - June

### Document history

Version	Date	Author	Status / Change
01	23-Jun-05	Simon Inger – Scholarly Information Strategies	Issued version

Please send suggestions or corrections for these notes to: Simon Inger – [simon@scholinfo.com](mailto:simon@scholinfo.com)

Project Report - June .....	1
1 PUBLISHER OVERVIEW .....	2
2 PUBLISHER CAPABILITIES .....	2
3 PROJECT RECOMMENDATIONS .....	3

	<b>Project Note</b>	Version 01	Status: Issued	Date 23-Jun-05
	Project: JCLD Pilot Project in Anticipation of E-Journals		Topic: Project Report - June	

This report has been created by Simon Inger of Scholarly Information Strategies Ltd, working as a consultant to the British Library.

## 1 PUBLISHER OVERVIEW

All twenty-three volunteering organisations have now been visited so far. Discussions with the publishers included a project overview, approximate timeline, data formats and data submission methods.

There are 205 titles included in the project.

The publishers in the project are a good cross section of the UK journals publishing sector. They include publishers large and small, commercial and not-for-profit with a wide variety of subject matter. In addition they have many different production and data distribution methodologies and produce data to a wide variety of formats.

Publishers have welcomed the project; archival of their e-journals is seen as an important function.

## 2 PUBLISHER CAPABILITIES

Most publishers would prefer to submit full text XML or SGML. Of these publishers, almost all considered the XML as the published form. In addition it was felt that their XML was a much better candidate for long-term preservation than PDF. Those XML feeds normally also include PDF files.


The remaining publishers would be supplying PDF files with associated metadata. In many cases this was because although the publishers produced SGML or XML full text for their web sites, they had no method by which such data could be harvested for third party distribution – yet. Other distribution partners tend to receive PDF/SGML feeds and so this is all that the publishers had built.

The great majority will provide XML or SGML headers largely to NLM or SSH2 DTDs. Two publishers have no SGML headers.

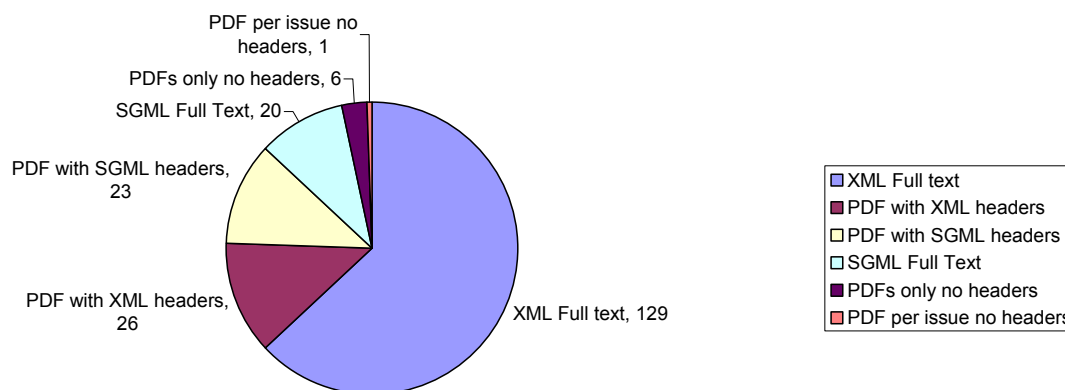
There was very little standardisation amongst full text DTDs.

The most popular DTD amongst publishers was the NLM DTD, in its SGML or XML variants.

In the market as a whole there appears to be greatest support for the NLM DTD. Other archival projects, notably Portico and LoC are likely to adopt it. Highwire is reported to be dropping its DTD in favour of the NLM DTD.

	<b>Project Note</b>	Version 01	Status: Issued	Date 23-Jun-05
	Project: JCLD Pilot Project in Anticipation of E-Journals		Topic: Project Report - June	

### Numbers of Journal Titles in e-journal pilot by Data Format



## 3 PROJECT RECOMMENDATIONS

A project meeting held 8<sup>th</sup> June, hosted by the BL and involving representatives from National Library of Scotland, Oxford and Cambridge agreed the following recommendations:

For preservation purposes, we will ask publishers that have XML full text to provide it to their own DTD. (Alternatively it could be supplied to the NLM DTD if this is a lossless transition)

For preservation of the presentation, we will ask publishers for PDFs, in the best quality that they have.

In the absence of full-text XML (or SGML) we will ask for SGML or XML headers to accompany the PDFs.

Where a publisher is unable to provide any tagged information, the project will look at other solutions including:

- Manual tagging
- Automated tagging from PDFs, such as that deployed by Ingenta, Atypon and Highwire.

Data submissions will be almost exclusively by FTP. It would be preferred if this could be a polling system – it gives greater guarantees of data collection. The project will look at OAI PMH as a basis for the polling.

Data may be submitted or collected from publishers, their data management partners (e.g. typesetters) or from their e-journal hosts.

Amongst the publishers taking part in the project, there were none that thought that they would benefit from the availability of a data submission tool.