SACAP Cooperative Collection Development Workshop
Background Document: OCLC Profile of South Asian Monographs
5/4/2010

The following charts attempt to set the scene for the proposed cooperative collection development workshop for SACAP. Based on some preliminary analysis of OCLC statistics, we hope to present the “big picture” of how publications from South Asia are represented in our collective institutions.

First, a caveat: The data behind this assessment comes from OCLC’s WorldCat database. It is acknowledged this is a far from perfect source from which to draw meaningful conclusions. Duplicate titles, incorrect holdings abound, based on anecdotal accounts from various institutions. Nonetheless, it remains the best available set of information from which to study our aggregate collection.

Collecting of monographs by SACAP institutions has remained largely steady over the past 50 years. Meanwhile, titles reported by all institutions in OCLC have grown significantly over time, suggesting that more institutions are collecting imprints from South Asia, from other sources, and possibly that SACAP is not keeping pace with the extent of publishing in South Asia. Based on some sampling (and factoring in duplication of records for titles, editions), it is reasonable to assume more than 100,000 titles from South Asia published in the past 25 years are not held by the Library of Congress or any other SACAP participant.

By language, English publications are far more readily accessible to scholars, while collections of publications in less commonly taught languages remain small. The number of titles collected in regional languages over the last 10 years has remained steady, even increased slightly, but at a far smaller scale than the major collecting languages.

The data suggests that titles from India are more commonly held among libraries than publications from other South Asian Countries. Any cooperative collection arrangement might do well to target commonly collected profiles from India above those from other countries.

We welcome your own interpretation of the data and can share the statistical tables and search strategies on request.

James Simon
Director, Global Resources Network
Center for Research Libraries

Source Data:
SACAP Titles by country - WorldCat searches 5/1 to 5/2/2010
SACAP Titles by language – WorldCat searches 4/28 – 4/29/2010
SACAP Overlap – WorldCat searches 4/29/2010
The chart above shows the growth of South Asian publications by date of publication, as held by all OCLC institutions, SACAP participants (excluding LC, CRL, and the National Libraries of Medicine and Agriculture). It also shows titles held by the Library of Congress as a base of comparison.

The data suggests collecting of monographs by SACAP institutions has remained largely steady for the past 50 years (± 10,000 titles). Meanwhile, titles reported by all institutions in OCLC have grown significantly over time, suggesting that more institutions are collecting imprints from South Asia, and possibly that SACAP is not keeping pace with the extent of publishing in South Asia.1

---

1 It is difficult to conduct a gap analysis of titles not held by SACAP due to the number of duplicate records contained in OCLC WorldCat.
These charts break down titles reported held by SACAP libraries (Non-LC, CRL, NAL, NLM) by country. India is represented separately due to scale.

The data reinforces the relatively consistent collecting of monographs across countries. Monographs from India hover between 32,000 and 39,000 titles per five-year period. Pakistan fluctuates ± 2,000 titles (30%) over time, while the remaining countries show steady (though modest) growth from 1960 to present.

The reason for the overall drop in titles for the period 2005-2009 is not certain, perhaps due to processing backlogs at SACAP institutions, an actual drop in collecting levels, or both.
75% of all titles recorded are in the most commonly-collected languages of English, Hindi, Urdu, Bengali and Tamil.
Figures 5 – 9 – Overlap in SACAP monograph collections, by country

Figure 5: India

![India copies held, 2000-2009]

Figure 6: Pakistan

![Pakistan copies held, 2000-2009]

The charts represented here (and on the next page) attempt to demonstrate the relative overlap in SACAP holdings from South Asia. This assessment was performed for each country for the years 2000-2009.

Note that this overlap study does not represent uniquely held titles. To derive a reasonable sample of SACAP-provided titles, the author searched for monographs held by Library of Congress and at least one other SACAP participant. The charts break down relative percentage of holdings by the number of holding libraries in OCLC (2-4 libraries, 5-9 etc). 

For India, roughly 50% of all SACAP titles are held by 9 or fewer libraries. For Pakistan, the percentage is higher than India, closer to 60-70%. For Sri Lanka and Bangladesh, the percentage fluctuates between 50-60%. Nepal has the highest percentage, at 70% or greater.

In the lowest category (2-4 libraries), Sri Lanka, Bangladesh, and Nepal retain high percentages of uniqueness (around 30%). Pakistan fluctuates between 10% and 30%. For India, approximately 20% of SACAP holdings are held by 2-4 libraries.

Overall, the percentages of relative uniqueness grow for 2007-2009, perhaps due to the aforementioned processing backlogs. Comparable studies of holdings over time performed by the CRL Global Resources Network support this hypothesis.

The data suggests that titles from India are more commonly held among libraries than publications from other South Asian Countries. Any cooperative collection arrangement might do well to target commonly collected profiles from India above other countries.

---

This study did not attempt to restrict the number of holding libraries to SACAP institutions only. Therefore, numbers of holding libraries represent all institutions within OCLC.
Figure 7: Sri Lanka

Sri Lanka copies held, 2000-2009

Figure 8: Bangladesh

Bangladesh copies held, 2000-2009

Figure 9: Nepal

Nepal copies held, 2000-2009