

# Digital Text Projects in Eastern Europe: Promoting International Interoperability

**Miranda Remnek** ([mremnek@uiuc.edu](mailto:mremnek@uiuc.edu))

Slavic & Eastern European Library  
University of Illinois at Urbana-Champaign

The widely divergent languages and cultures of Russia, Eastern Europe & Eurasia (hereafter Eastern Europe), present a rich terrain for digital scholarship, but owing to a number of factors—including poorly-endowed or mutually-incompatible infrastructures, language barriers, and different approaches to the construction of metadata—the corpora already in existence and those in development are often poorly known elsewhere, and in many cases are hidden from international resource discovery and data sharing.

In view of this situation, it was clear to the Digital Projects Subcommittee of the *American Association for the Advancement of Slavic Studies* (<http://www.aaass.org>) that improved documentation would constitute an important step toward the promotion of knowledge sharing and interoperability.<sup>1</sup> The Slavic & East European Library at UIUC is therefore building an international *Inventory of Slavic, East European and Eurasian Digital Projects* (<http://www.library.uiuc.edu/spx/inventory>). The *Inventory* already describes at least 360 collections in 110 projects, and we also foresee web submissions from international partners to promote collaborative registration. But even now the *Inventory* is heavily used; recent statistics indicate that in the period July 2005-June 2006 it received around 25,000 hits. It has also become an OAI Data Provider, and although inventory contents (unlike the digital collections they record) are sometimes not reflected in aggregated search services like OAIster (<http://oaister.umdl.umich.edu/o/oaister/>), the *Inventory's* records have indeed been harvested and now display in OAIster searches—likely to improve dramatically the visibility of substantive digital projects in the Slavic field.

But the project goals go far beyond the compilation of a mere inventory. Besides the implementation of intensive content development, promotion of metadata standardization, expanded reference assistance and interactive user options, the *Inventory* will focus on an expansion of the OAI data provider system and item-level harvesting. (It also hopes to explore customized delivery services and an archival framework for at-risk

collections in the Slavic field, currently being explored through an inter-institutional grant proposal).

Nevertheless, the project team acknowledges that metadata practices in Eastern Europe are still very diverse, and far from standardized. At one end of the scale, where relatively simple encoding is involved, even schemes like Dublin Core are poorly recognized. This makes the implementation of relatively well-established data-sharing protocols like OAI-PMH (*Open Archives Initiative: Protocol for Metadata Harvesting*) little known in Eastern Europe. A registry of OAI data providers at UIUC demonstrates that while OAI data providers are well established in the US and Western Europe, they are much less widely encountered further East.<sup>2</sup> True, the implementation of OAI data interchange is not necessarily well standardized in countries with far more providers.<sup>3</sup> Furthermore, the whole question of the efficacy of OAI-PMH as a metadata transfer protocol is still somewhat open to question.<sup>4</sup> Nevertheless, its power is generally well-recognized, and its further penetration into Eastern Europe is certainly a desirable goal.

At the other end of the scale, the level of deep encoding is even less widely standardized in Eastern Europe. The decade-long Institute of World Literature project in Moscow known as FEB (*Fundamental'naiia elektronaiia biblioteka*) (<http://www.feb-web.ru/>) has an enviable collection of over 50,000 Russian literature texts, all heavily encoded in SGML—but not according to the *Text Encoding Initiative* (TEI) Guidelines. True, the TEI is not a standard, but interoperability with such an impressive collection would be highly desirable. Other such examples are available. To be sure, there are exceptions: the recent TEI meeting in Sofia, Bulgaria revealed the extent to which TEI projects are beginning to gain root in South Eastern Europe.<sup>5</sup> There are also similar developments in a more abstract sense. The TEI Guidelines have been translated into Russian,<sup>6</sup> and efforts are also being made to promote the further internationalization of the Guidelines, and to ensure their availability in a greater number of languages, including Bulgarian.

Yet even when East European scholarly projects adopt the TEI, standardized approaches to the conversion of metadata to OAI-compliant formats are also at issue, and there is evidently much ground still to be covered. Hence, the object of this paper will be to:

1. Design and conduct a survey of metadata practices in a number of East European digital centers (including existing UIUC partners) that are sponsored both by institutions and by individual faculty teams;
2. Produce an up-to-date analysis of their awareness, evaluation and observance of international metadata standards;
3. Identify problems and practices preventing their implementation;

4. Suggest ways in which steps can be taken to ameliorate this situation, including crosswalks and other technical procedures.

The results of this exercise will not only assist in a practical sense the development of the *Inventory of Slavic, East European and Eurasian Digital Projects* (the only registry of substantive, scholarly East European digital initiatives), but will also result in better awareness and information-sharing among Slavic digital practitioners. This is especially vital as the institutional repository movement gains ground.<sup>7</sup> Experience shows, particularly in West Europe, that the issue of standardized metadata is paramount, and that subject repositories are becoming of even greater interest since they tend to be populated by scholars working on their own projects and producing more substantive metadata. In this environment, East European scholars should not be excluded from the pool of informed experience.

---

But in Eastern Europe only 2 DSpace sites were listed (in Russia). This lack of awareness was substantiated at an e-text conference in Eastern Russia in July 2006 entitled *Modern Informational Technologies and Written Heritage: From Ancient Manuscripts to Electronic Texts* at which I gave a presentation entitled "Promoting the TEI in scholarly communities: OAI harvesting, digital repositories." Conference attendees described many sophisticated projects, but there was little familiarity with OAI or digital repository software used in the West. One scholar was familiar with these technologies, but in a presentation entitled: "A Computer System for the Creation and Maintenance of Electronic Collections of Ancient Texts," V. S. Iuzhikov (Kazan State University) wrote: "For the creation and development of electronic libraries there are several systems. Among the best known are Greenstone and DSpace. But they are oriented mostly toward text libraries with small number of illustrations, which are poorly suited for the collection of old printed works. Therefore it is less time-consuming and gives better result to build a local system." Given the specialized nature of his material, this conclusion was understandable. Yet the general lack of East European involvement with these international standards and methodologies remains true.

1. Other points in the group's charge include: (2) fostering informed participation in future initiatives; (3-4) increasing digital information and training opportunities for Slavic librarians; (5) complementing national efforts to establish digital repositories. See <<http://www.library.uiuc.edu/spx/BnD/DigPro.htm#charge>>
2. See <<http://gita.grainger.uiuc.edu/registry/ListTLDs.asp>>. U.S. providers are proliferating: in the past year the number of "edu" domains increased from 265 to 332, and "org" domains from 149 to 182. Likewise in Western Europe: U.K. DPs went from 89 to 114, Germany from 73 to 99. But in Eastern Europe it is different; the totals for Hungary (4), Czech Republic (2) and Slovenia (2) stayed the same this year, and Russia increased only slightly, from 3 to 4. Only Poland increased significantly, from 4 to 18.
3. As reported at the Oct. 2005 OAI workshop in Geneva, the German Initiative for Networked Information (DINI) has installed a certificate system in order to bring OAI providers into greater standardization.
4. The Open Archives Initiative OAI-Implementers group announced on Nov. 3 2005 that it is "studying the effectiveness of OAI and some other related methods of creating interoperable online libraries" and has "posted a questionnaire 'Survey on Common Interface Frameworks for Online Libraries' on the web."
5. See Milena Dobrova, "TEI in South-Eastern Europe: Experience and Prospects," TEI Members Meeting, Sofia, Bulgaria, October 28, 2005.
6. See <[http://www.tei-c.org.uk/Lite/teiu5\\_rtf](http://www.tei-c.org.uk/Lite/teiu5_rtf)>.
7. Here again the number of East European repositories lags far behind. In July 2006 there were 144 DSpace instances worldwide, and less than half were in the U.S. and U.K. (62), indicating that many non-English-speaking countries (like Japan) are aware of and attracted to this increasingly popular open source software.