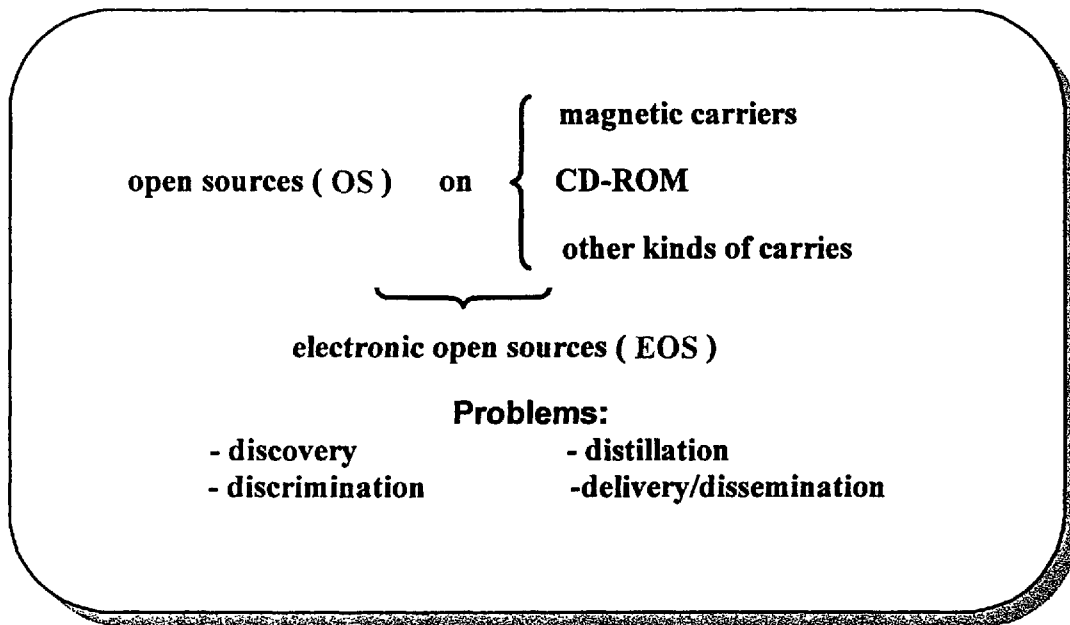


## The Russian Viewpoint on Electronic Open Source Technologies

Major General Vladimir Budzko

OSS21, 17 May 2000



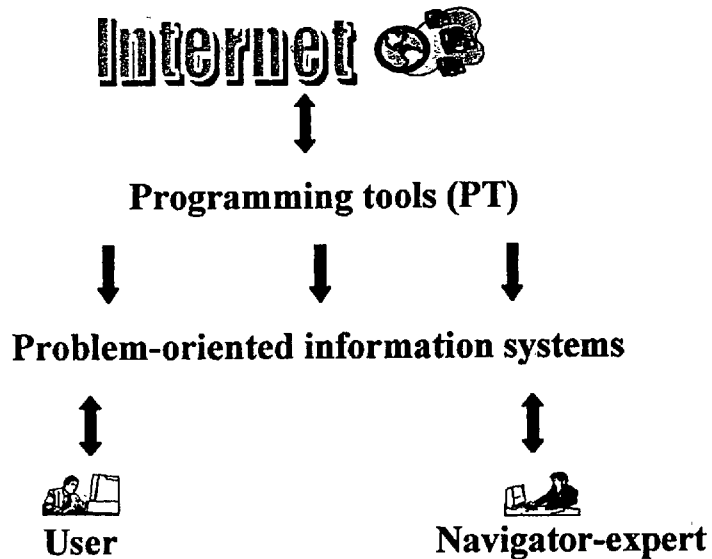
For this presentation, I define electronic open sources as those that are contained on the Internet, including commercial fee-for-service sources accessible through the Internet; and those that are contained in other electronic media including traditional magnetic carriers, CD-ROM, ZIP disks, DVD, and other emerging remote storage and wireless alternatives.

This complicated and constantly changing electronic "terrain" represents a very significant problem for the information technology managers responsible for each of the following four functions defined by Robert Steele as essential to open source exploitation:

- **Discovery:** how do we find this information, most of which has not been properly indexed in some sort of universal directory?

- **Discrimination:** in the absence of substantive editorial quality control, meta-tags, independent reviews, or other forms of value-added discrimination, how do we evaluate each item, both in isolation and in relation to all the other alternative electronic sources?
- **Distillation:** once we have solved the first two problems, how do we then weight, cluster, matrix, or otherwise organize all of this information in different formats and different electronic media, so that we can distil the information to its essence while retaining its heritage information and links to the original raw material?
- **Dissemination:** finally, when we are satisfied that the electronic open sources have been properly processed, how do we make the information accessible to the wide variety of constantly changing end-users, most of whom have very low bandwidth and older generation Internet tools?

## New information technologies and the Internet



In our view, the Internet offers great potential as the common foundation for executing each of these important tasks, but the Internet will not reach its full potential until certain programming tools and problem-oriented information systems have been developed that can effectively link electronic open sources, expert human navigators, and human all-source analysts and human decision-makers.

**PT functions:**

- 1. Creation and maintenance of the problem-oriented databases.**
- 2. Problem Structuring and the Internet Space cataloguing on the given subjects.**
- 3. Automatic selection of necessary information.**

The most difficult of all problems—leaving aside the issue of discovery—is data entry and cataloging. The multi-media database must be able to accommodate all kinds of electronic open sources, including HTML pages, digitized text, audio and video materials, etcetera. This complex database must be constantly indexed, and must allow for rapid and easy integration of new information as well as new links or categorizations based on human intervention.

**PT functions:**

- 4. Automatic information delivery to users .**
- 5. Search.**
- 6. Actualization of new sources.**

One solution to this problem, but it is still labor-intensive because the automated tools are not yet mature, is to have problem-oriented portals to the Internet that are also used as portals to internal databases. Such portals must combine expert humans whose judgements are trusted, with a long-term institutional commitment to discovering, indexing, and maintaining all forms of electronic open sources pertinent to the specific problem to which the portal is dedicated.

Such portals, each serving the needs of a distinct community of expert users, will be very big and bulky and may not appear useful or understandable to experts from other communities. Therefore we believe there needs to be a generic portal that serves as the umbrella or bridge from generic users into the specialty databases, in this way tailoring what parts of the expert portals might be visible to the generic user.

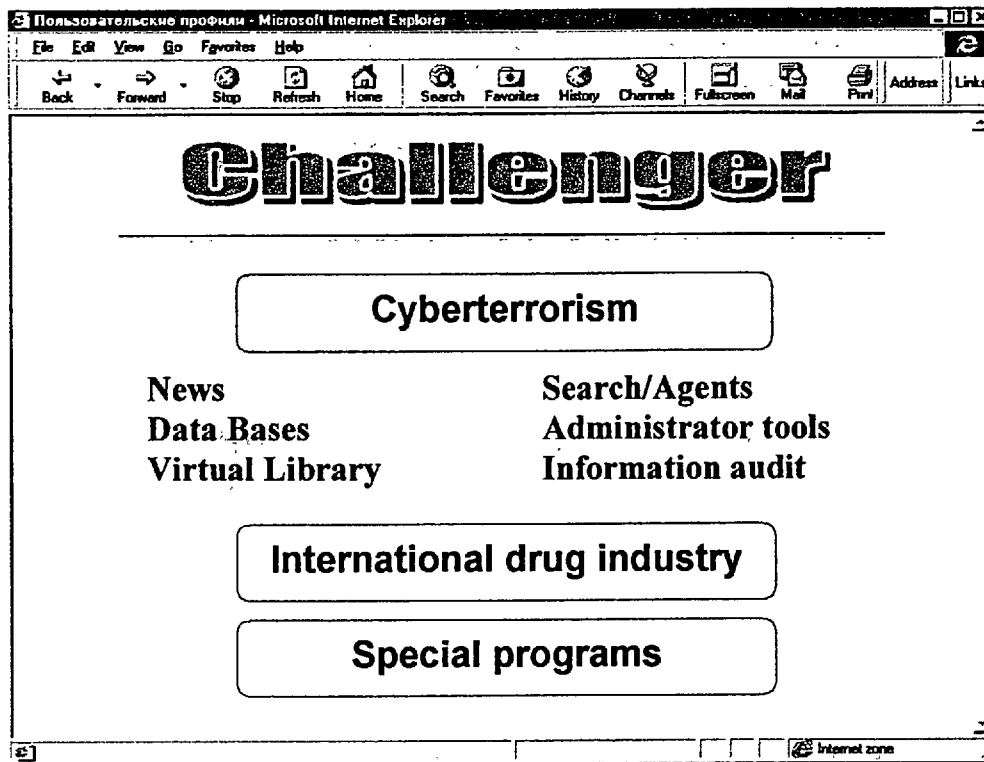
**PT functions:**

- 7. Securing information and audit.**
- 8. Creation of user thematic profiles.**
- 9. Administration.**

Information security and information audit and information reliability are a very important part of the electronic open source information process, and we do not see enough attention to this important aspect.

This is a complex area that includes identification of users, differentiation of authorities, controls on access, anti-virus controls, firewall features, etc.

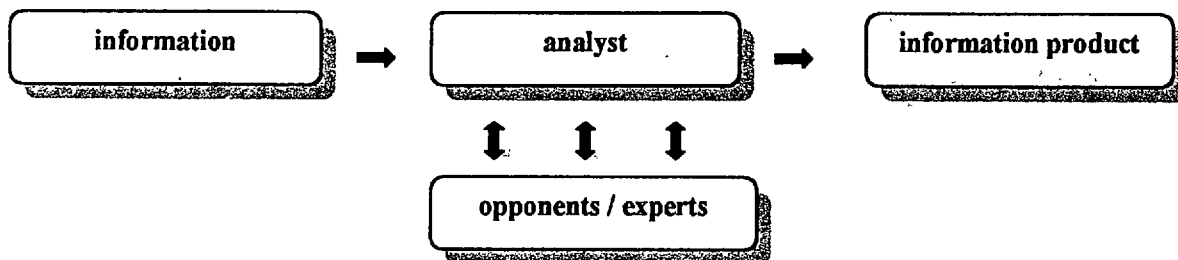
However, it must also include monitoring for user satisfaction, system efficiencies in answering questions, improved means of structuring and linking information, etc.



In Russia we are testing a new Internet resources monitoring shell, CHALLENGER, that had been created by a Russian commercial organization. This new product is a good example of how the Internet can be used to create a new style of work, but I would also emphasize that the Internet is not useful by itself—we must fully integrated programming tools and problem-specific programs that are expert-friendly, for the Internet to reach its full potential.

**Technical problems:**

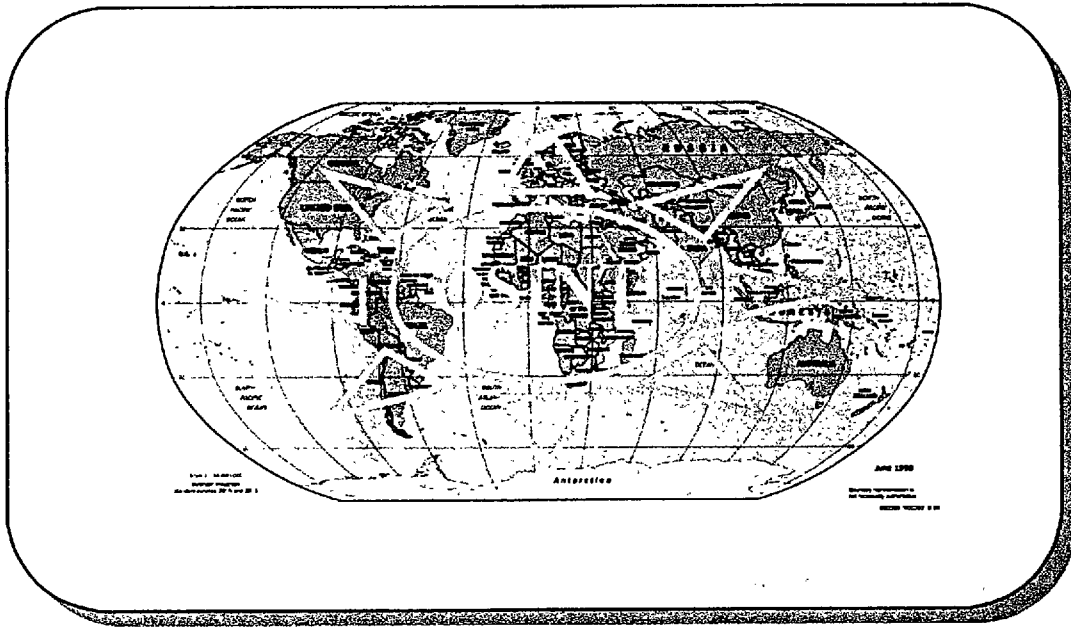
- **filtering large dataflows**
- **automatic translation**

**Non-technical problems:**

I will conclude by noting that there are still two serious problems that we have not solved completely in Russia and we do not see solutions available elsewhere:

- Effective automated filtering of large information flows, including mass media as well as internal data and contractor-generated research reports, is still not a reality.
- High-quality machine translation that can keep the sense of the source text is still not available for the majority of the languages of the world. The Russian commercial product **PROMT** (known in Europe as **REVERSO**) is the leader in Russia and provides a “good enough” output for certain major languages, including English (but as long as well-chosen dictionary is used). It is fully compatible with the Internet and with Microsoft documents. Of course we also need to accelerate funding for machine translation for Indonesian, Thai, Vietnamese, and many many other languages of importance to the future of the world.





One final note: I agree with Robert Steele's newest message, that such solutions are unaffordable by any single organization. I even agree with his larger statement, that such solutions are unaffordable and unachievable by any single *government*. We *must* find ways to work together in order to establish common requirements, to share the cost of developing new generic information handling capabilities, and to share the cost of open source data discovery, discrimination, distillation, and dissemination in those areas where it makes sense to share.

I am here because we are actively interested in international cooperation in open source intelligence activities.

# OSS 21 PRIMER Essential Elements of Information Joint Planning, Operations Other Than War and Open Source Intelligence - Link Page

[Previous](#)      [OSINFCEN STAFFING](#)

[Next](#)      [Electronic open sources: Technology of application](#)

[Return to Electronic Index Page](#)