

initiative B@BEL

www.unesco.org/webworld/babel



United Nations Educational,
Scientific and Cultural Organization

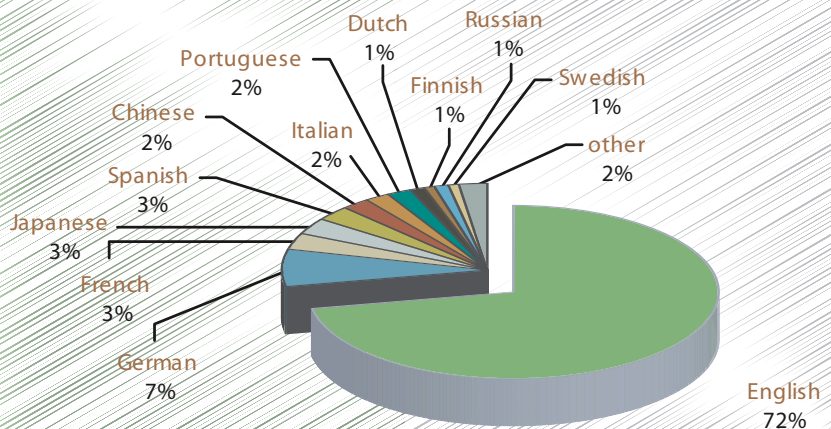
In today's emerging knowledge society, information is a key resource for decision-making, empowerment, social participation and human development. So increasingly the ability to access and influence knowledge / information on global information networks, such as the Internet, will influence one's ability to participate in society and determine one's quality of life. Language, our primary means for communication, is clearly crucial to this process. But languages are more than communication tools, they are vehicles of value systems, cultural expressions, an important factor in the identity of groups and individuals. Language diversity is an essential part of the living heritage of humanity that must be nurtured and promoted.

Initiative B@bel is contributing to UNESCO's goal of building knowledge societies, fostering universal information access and information lead development, through a three-fold strategy incorporating actions in the areas of policy, awareness raising and the implementation of innovative pilot projects. These activities carried out in collaboration with partners, seek to support the development of multilingual content and tools on the global information networks and protect endangered languages and their legacy of traditional knowledge. Initiative B@bel is a practical example of the application of strategies elaborated in such internationally adopted UNESCO standard-setting instruments as the Recommendation concerning the promotion and use of multilingualism and universal access to cyberspace.

While many countries have 2 or even 3 official languages, the 1993 Constitution of the Republic of South Africa recognizes 11 official languages.

Most alphabets contain 20-30 symbols, but the relative complexity of the sound systems of different languages leads to alphabets of varying size. The smallest alphabet seems to be that of Rotokas, used in the Solomon Islands, with 11 letters. The largest is perhaps that of Khmer with 74 letters.

An Estimate of Language content on the Web



Source : Proportions of languages on the web in a random sample of web pages. O'Niell, Lavoie and Bennet, 2003.

Guidelines and Studies Supporting Policy and Awareness Raising Initiatives

Education in a Multilingual World

This policy paper was developed by UNESCO in conjunction with a group of prominent international experts in the fields of cultural diversity and education. In addition to amalgamating important regional perspectives on key challenges and issues it also recalls and highlights existing international

agreements and commitments by governments to support linguistic pluralism and multilingual education in their territories. The document is available in Arabic, Chinese, English, French, Russian and Spanish.

National Terminology Planning Guidelines

A language community, which has not developed scientific and technical terminologies for their language, is inevitably forced to use another, perhaps foreign, language to express these concepts. This process reinforces language and knowledge divides and can contribute to a decline in the perceived importance, marginalization and exclusion of local languages from the wider knowledge society.

Terminology planning has an important role in addressing these challenges, supporting innovation, knowledge sharing, educational, economic activities and the wider developmental process. By providing policy makers, particularly those in developing countries, with a clearer understanding of this often overlooked area the guidelines seek to enhance the role of local languages in the knowledge society.

Thematic Edition of the International Journal on Multicultural Societies "Language on the Internet"

The Internet is a prime example of how ICT affects patterns of language use. These changes can have a number of far-ranging social, economic and other effects so public language policies may provide useful responses to these new challenges and opportunities. The research carried out breaks new ground by collecting and analyzing empirical data on patterns of language use on the Internet in the African, Arab, Asian and European regions.



Sample Script Implementations

While a system may be technically capable of handling multiple languages/scripts it is necessary to provide the basic building blocks (fonts, keyboards etc) that allow the entry and display of data in these scripts. In parallel with our multilingual tool

development we have developed script implementations for Myanmar and pan-African languages, allowing these tools to be used by the language communities.

Guidelines for Development of Writing System Support

This document presents an overview of key issues, activities and strategies for supporting the development of writing systems in the electronic environment.

Complex Script Data Exchange Format

For successful information creation and dissemination, particularly for complex script data, a standard data exchange format is needed.

Intellectual Property Rights and Writing Systems

The development of adequate computer writing system implementations and data handling for complex scripts is a challenging undertaking requiring various types of expertise and investment. Consequently, intellectual property rights (IPRs) may pose a hindrance to the refinement, distribution and

use of these systems by disadvantaged language communities. This paper explores models and policies that stimulate the development and distribution of writing systems while protecting commercial interests and IPRs and fostering content creation in many lesser-used languages.


Handbook for Linguistic Data-Collection by Non-Linguists

Traditionally linguists have visited various parts of the world to “discover” study and record “new” languages. The limited number of linguists, rapid acceleration in the disappearance of endangered languages coupled with other challenges make this approach inadequate. This project however seeks to empower interested endangered language communities to preserve their language heritage by providing them with appropriate methodologies for documenting it. This will serve to halt language loss, allow more advanced linguistic study of these languages as well as support the eventual development and dissemination of pedagogical material and indigenous local content both in print and digital formats.



It is estimated that about 20% of the world's languages have no written form. The emergence of multimedia technology, managing of digitised voice communication and innovations such as voice recognition and translation systems present new possibilities for the preservation of the rich heritage of these languages in today's digital environment.

Over 50% of the world's 6000 languages are endangered and 90% of the world's languages are not substantially represented on the Internet.



Pilot Projects, Online Resources and Multilingual Tools

Report on Multilingualism on the Internet

How many languages are present on the web? How many pages exist in each of these languages? What are the challenges to measuring linguistic diversity on the Internet and how can the technology used for this purpose be improved? What forces are shaping the linguistic fabric of the Internet and how is this

changing over time? This paper developed in cooperation with UNESCO Institute for Statistics and which has been incorporated in the World Summit on the Information Society Publication Series seeks to provide us with a better understand of these and other related issues.

The Multilingual Content Management Platform (MLCM)

In cooperation with Infoterm, the International Centre for terminology (www.infoterm.org) in Vienna, Initiative B@bel has developed a multilingual information management system based on UNESCO's CDS/ISIS platform ensuring compatibility and data reuse. The software which incorporates functionalities such as bar-code

scanning, compatibility with Microsoft's Office suite is also capable of carrying out multilingual searches within local area network and on the Internet. The MLCM which targets the information management needs of SMEs contains extensive and extendable factual and terminological databases.

SILA – A Multilingual Web Browser

The web is an important forum for information dissemination exchange and interaction but many language communities are excluded as web tools for their scripts are unavailable. By integrating SIL's Graphite technology for displaying complex writing systems with the Mozilla Internet application suite

(Browser, Email, HTML Composer, IRC Chat and Address Book) and developing support fonts Initiative B@bel is creating the tools that enable marginalized language communities to benefit and interact on the web. In the first phase support for Burmese and African languages is being provided.

Modified version of WorldPad

WorldPad is a simple text editor and data input tool that can seamlessly handle complex (non-roman)

scripts using the Graphite rendering engine developed by SIL International.

Graphite Edit Control

A beta Graphite Edit Control for data input application, using SIL International's Graphite engine and software development kit. This tool is particularly useful for applications where text blocks contain several scripts, (for example bibliographic

entries containing the names of a book as it appears in various translated editions are encountered) and supports both left to right and right to left (e.g. Arabic, Hebrew) scripts.



At present about 90% of internet content is available in only 12 languages.

Multilingual Cultural Heritage Knowledgebase

This project combines thematic knowledge on cultural policies with the Universal Network Language translation system developed by the UNDL Foundation. On completion this knowledge base will provide a platform including expandable built-in dictionaries and specialized terminologies

and a cultural policy database accessible in 6 languages (Arabic, English, French, Italian, Russian, Spanish). As the system is modular and contains a user development kit, support for other languages and fields of knowledge can be integrated creating a multilingual encyclopaedia of knowledge.

Internet Sound Record Library of Endangered Caucasian Languages

There are over 40 endangered languages in the Caucasian Region many with unique and distinctive linguistic features. As part of its efforts to protect languages and their associated rich body of knowledge and traditions an Internet accessible sound record library of endangered languages of the

Caucasian Region is being developed. In addition to providing greater awareness to the urgent plight of these languages this resource will also allow linguist to further study these languages and contribute to preservation and perhaps revival of the linguistic and cultural heritage for future generations.

Abkhazian-Georgian Dictionary Project

Georgia has a wealth of cultural and natural resources but the ongoing civil conflicts over the past decade have exerted a heavy toll on the cultural, social, physical and economic fabric. In support of the peace building process a cooperative effort between academic institutions has been initiated to develop a bilingual (Abkhazian-Georgian) print and

electronic dictionary. In addition to developing multilingual resources and by building of bridges for communication and opening/creating pathways for dialogue, commerce and other information exchanges, this project is seeking to contribute to a lasting peace.



two-volume
ABKHAZIAN-GEORGIAN
general purpose dictionary



Ethiopic Computer Standards



Computer users wishing to compute in non-roman scripts may face practical challenges such as incompatible information encoding systems and a lack of standards for keyboard layouts and transliteration. As a result, information sharing between users may be difficult if not impossible and data entry and other processes may become inefficient, error-prone and costly thereby discouraging the use of local languages in the digital environment. In response to this challenge, Initiative B@bel, in concert with the government of Ethiopia

and other key partners from the international community, private sector and academia, conducted a study to assess and identify computational support needs for Ethiopic languages, and to develop standards for Ethiopic character encoding, keyboard layout and transliteration of Amharic into the Latin alphabet. On this basis an action plan for addressing these challenges and a working group have been established. Currently efforts to secure funding to implement these changes are underway.

```
with (navigator)
  if ((appName=="Netscape")
    &(parseInt(appVersion)==4)) {
```

```
<option value="2">
```

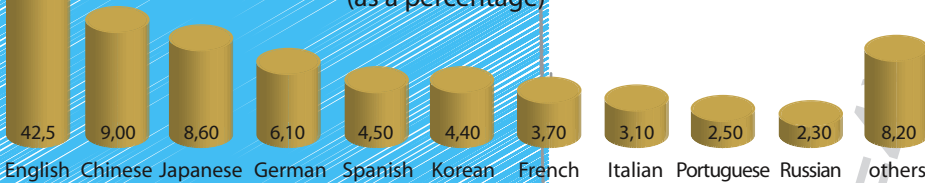
UNESCO's Atlas of Languages in Danger of Disappearing

Initiative B@bel is developing an online extension of this widely acclaimed print publication. This online product will provide new insights into the socio-linguistic and cultural life of the people who use these endangered languages. The initial focus will be to present information on Africa's endangered language communities.

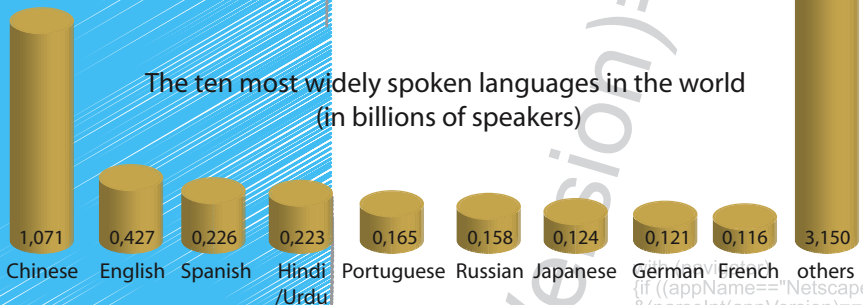
Over time, interactive multimedia functionalities along with GIS (geographical information systems) data management support will be incorporated. In addition to enhancing users' experience and promoting awareness, we hope to create a tool that can provide easily accessible and regularly updated information on the state of these languages.

on <http://www.unesco.org>.

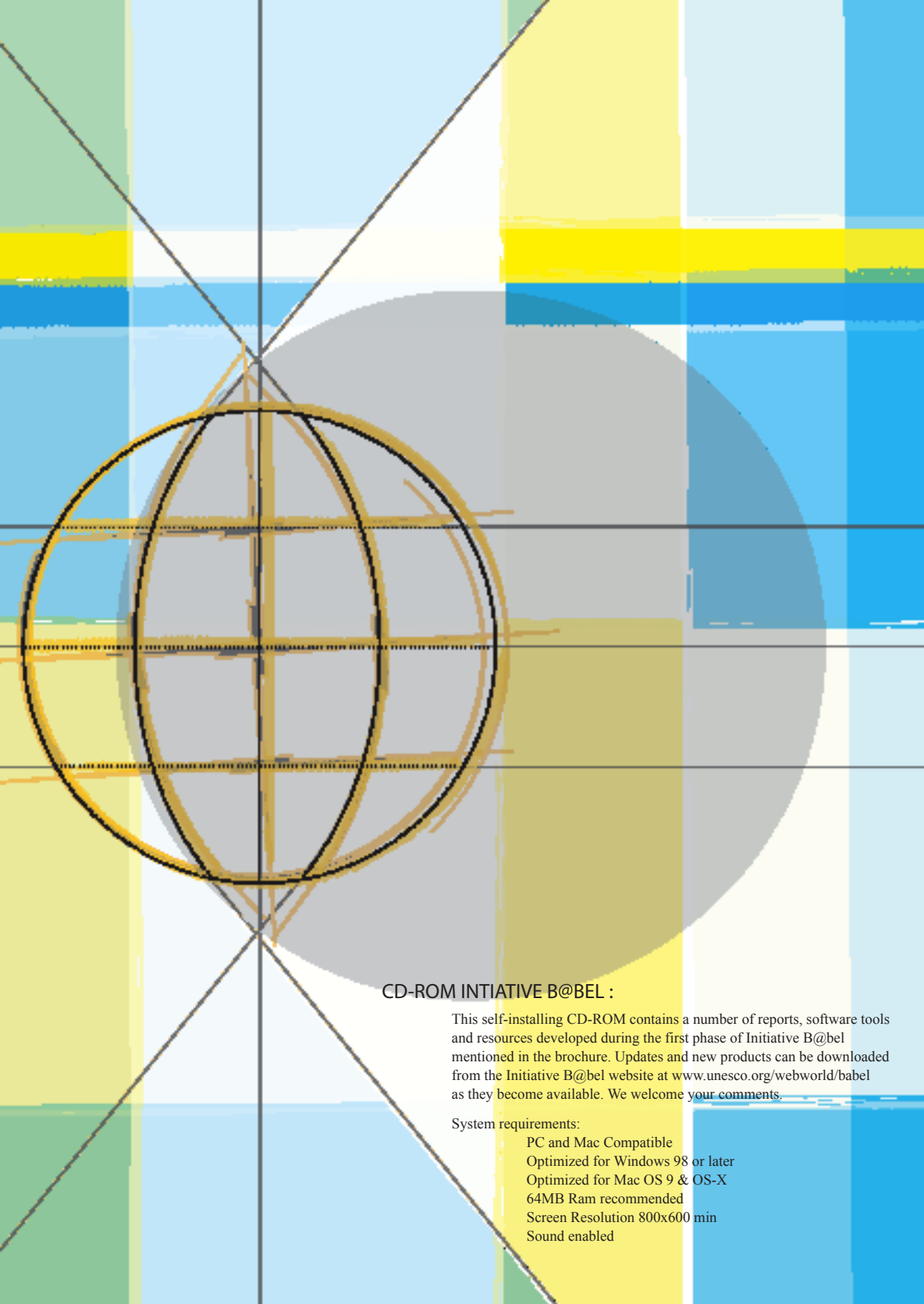
The biggest language populations on the Internet (as a percentage)



The ten most widely spoken languages in the world (in billions of speakers)



Source : The Cambridge Encyclopedia of Languages. D. Crystal, 1992. p289



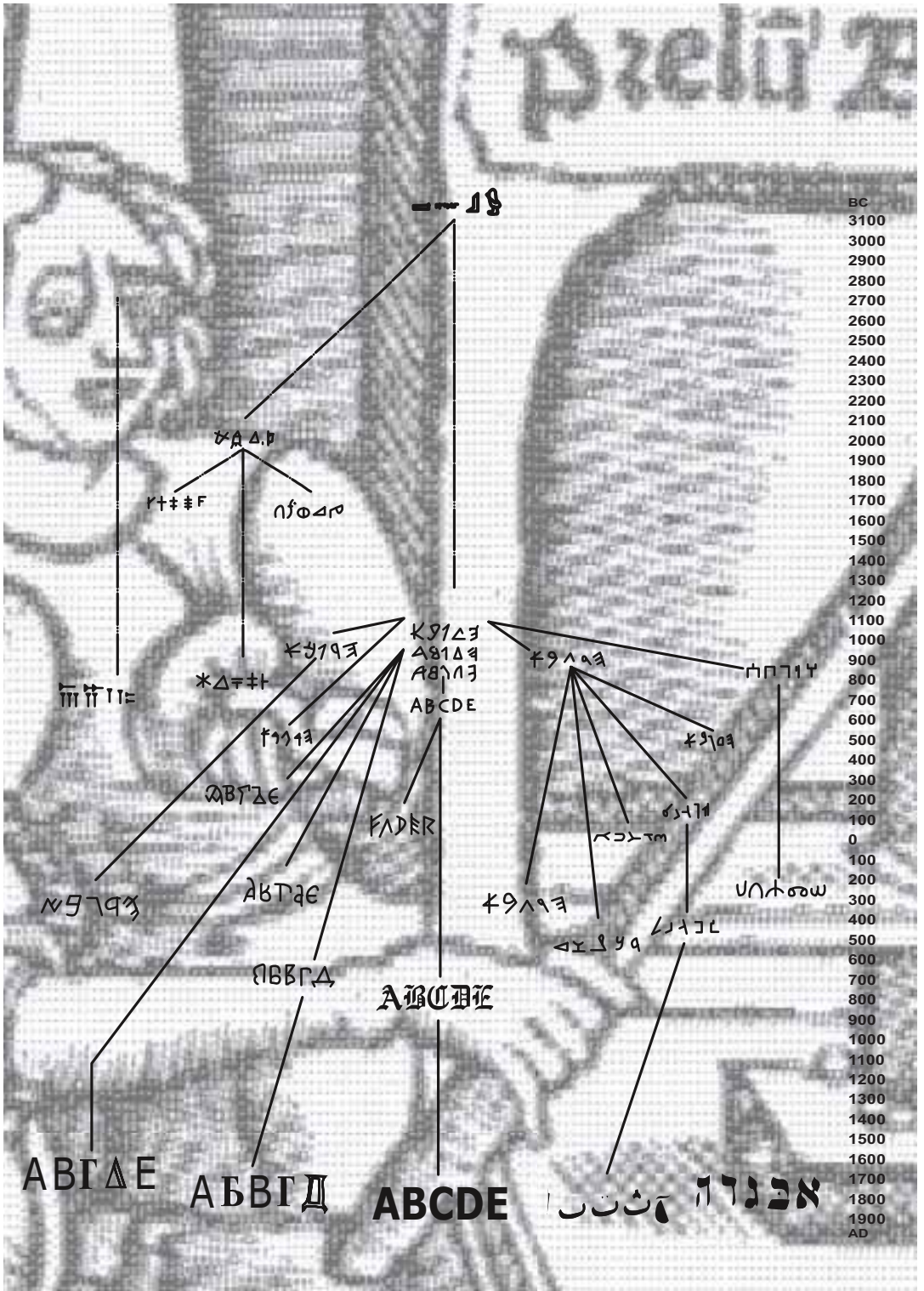
CD-ROM INITIATIVE B@BEL :

This self-installing CD-ROM contains a number of reports, software tools and resources developed during the first phase of Initiative B@bel mentioned in the brochure. Updates and new products can be downloaded from the Initiative B@bel website at www.unesco.org/webworld/babel as they become available. We welcome your comments.

System requirements:

- PC and Mac Compatible
- Optimized for Windows 98 or later
- Optimized for Mac OS 9 & OS-X
- 64MB Ram recommended
- Screen Resolution 800x600 min
- Sound enabled

'The world's alphabets over the centuries'



Source: The Cambridge Encyclopedia of Languages, by David Crystal ; Mary Evans Picture Library

Selected Upcoming Initiatives :

- Multilingual translation systems
- Online language learning
- Voice to Speech
- Unicode script encoding
- Font development and training font designers
- Language support for persons with disabilities



United Nations Educational
Scientific and Cultural Organization

References & Acknowledgements :

- The Cambridge Encyclopedia of Languages
- Mary Evans Picture Library, London, England
- Stockholm Challenge
- UNESCO Atlas of Languages in Danger of Disappearing (Edition 1,2)
- UNESCO Photobank
- Mr Victor Montviloff
- Lucia Dessart, Christina Human, Nathalie Da Fonseca,
Sabine Schaefer, Marina Forzani, Elena Iasyreva
- Mr Paul G. C. Hector

Project Partners and Supporters to date :

- Government of the Principality of Monaco
- Government of the Federal Democratic Republic of Ethiopia
- Georgian National Commission for UNESCO
- Swiss Agency for Development and Cooperation
/ICT4D/Global Knowledge Partnership
- African Academy of Languages
- UNESCO Institute for statistics
- Infoterm
- International Literacy Institute
- UNDL Foundation
- Ethiopia Computer Standards Association
- SIL International
- Aston University
- Matenadaran, Mashtots Institute
- University of California (Berkeley)



Themes of the design :

The brochure design incorporates elements such as colours, the earth and symbols of varied cultural origins that prepare the reader to explore multilingual communication. The central arrow leads the reader on a cyclical journey that retraces languages from their origins to today's digital media. Its cyclical nature implies a dynamic where notions of past and future fade to provide a perspective of a global village where all have the right to information. Characters from various scripts watermark each page and remind us of the diversity of writing systems and the need to protect, preserve and represent them ALL digitally. There are also other meaning for you to uncover on your own such as Gutenberg meeting ASCII code.....

Jean-Luc Seyfried Freelance Designer
www.bitspersecond.org

To obtain more information, or to become involved in our activities, please
contact us to the address below :

Communication and Information Sector Information Society Division

1, rue Miollis 75732 Paris Cedex 15, France
Tel. + 33 (0)1 45 68 43 62
Fax. + 33 (0)1 45 68 55 83
Site: www.unesco.org/webworld/babel
E-mail: Babel@unesco.org