SC22/WG20 N974

Programming for cultural diversity consensus-building

a Project Team report to CEN/ISSS

Project Team:

John Clews (project leader)
Keytempo Limited
8 Avenue Rd
Harrogate HG2 7PG
United Kingdom
cdiv@sesame.demon.co.uk

Håvard Hjulstad (editor)
Solfallsveien 31
NO-1430 Ås
Norway
havard@hjulstad.com

Finalization version
2002-08-14

Executive summary

The CDICT Project Team, which was set up by the Cultural Diversity Steering Group, began its work in mid-April 2002.

The Terms of Reference required a report setting out a programme of work for standardization in the area of cultural diversity and including proposals for where it should be carried out, and its dealings with other bodies on CDICT issues, and progression of current work under BC/CEN/97/26.

This report provides this, taking into account

- the PricewaterhouseCoopers' (PWC) report on standardization/consensus-building measures;
- the PWC Annex on CEN / TC 304 (*Information and communications technologies – European localization requirements*);
- the CEN / TC 304 reply and commentary;
- the list of current work under BC/CEN/97/26;
- Resolution CEN BT 12/2002 (June 2002), concerning the "Status and Secretariat of CEN / TC 304";
- Recommendations from the CEN / TC 304 Chair (June 2002) about possible future status of existing work in a more dormant CEN / TC 304.

The report gives an overview of a "Strategy for Standardization of Cultural Diversity in ICT systems" in chapter 1. Chapter 2 gives a review of the current situation, listing organizations and standards-relating activities in the field of Cultural Diversity in Information and Communication Technology (CDICT).

Chapter 3 looks more closely at the "Scope of Cultural Diversity in ICT", discussing various taxonomies in looking at CDICT. That of the eEurope Action Plan was chosen as the best fit for the future scope for standardization within Cultural Diversity in ICT systems.

The proposed work programme is described in chapter 4, structured according to the eEurope Action Plan:

1. A cheaper, faster, secure Internet
   a) [infrastructure] Cheaper and faster Internet access
   b) eResearch Faster Internet for researchers and students
   c) eSecurity Secure networks and smart cards
2. Investing in people and skills
   a) eLearning European youth into the digital age (eEducation)
   b) eWorking Working in the knowledge-based economy
   c) eAccessibility Participation for all in the knowledge-based economy
3. Stimulate the use of the Internet
   a) eCommerce Accelerating eCommerce
   b) eGovernment Government online: electronic access to public services
   c) eHealth Health online
   d) eContent European digital content for global networks
   e) eTransport Intelligent transport systems

The report proposes to add two elements to this list: (a) machine translation and (b) sound applications.

The Project Team's recommendations are found in chapter 5 of this report. 11 recommendations are listed.

An advisory group is proposed, for the purpose of reference in this report named "CDICT Advisory Group for Europe" (CAGE). The new CAGE would include among its responsibilities the activities of the current CDSG.

The Project Team recommends the endorsement of the proposal to make CEN / TC 304 dormant once certain projects have been finalized.

Chapter 6 sets up an initial work programme for CAGE. The work programme is closely linked to the eEurope Action Plan for 2005 and the eEurope Action Plan for 2002, with some additional areas of work. The work programme also discusses the continuation of unfinished work items stemming from CEN / TC 304.

Chapter 7 is a draft Business Plan for CAGE.

Table of content

0 Introduction

The Cultural Diversity Steering Group (CDSG) and the Cultural Diversity Project Team were established to assist CEN/ISSS to work out ways forward in relation to standardization and related aspects (including implementation and use) of Cultural Diversity in ICT (CDICT), which is important for CEN members, given the multinational and multilingual state of the CEN area.

The CDICT Project Team, which was set up by the CDSG, began its work in mid-April 2002, and reviewed standardization efforts and structures across standards groups concerned with Cultural Diversity in ICT. Interim recommendations are made on the basis of that initial review.

In a European context, CEN / TC 304 (*Information and communications technologies – European localization requirements*) started in 1992 with character sets and embarked in late 1997 on a new large work programme moving into many other areas of CDICT. Four of the work items were moved to CEN Workshops in 1998. Consensus was at times difficult to reach in the work on some work items. The PricewaterhouseCoopers report tried to evaluate this large programme.

A formal Technical Committee with Working Groups, and Workshops have been used within CEN/ISSS for this purpose, with varying assessments about the success of both systems, and some of the results.

Although overall CEN/ISSS Workshops have had generally favourable results, over a relatively short time period, and although it may be tempting to go for a “Workshops good, TCs bad” approach (to drastically oversimplify it to get the point across), there may be other useful approaches that can be adopted too.

During the course of this project, various global approaches have emerged in relation to standardization involving Cultural Diversity in ICT systems (CDICT):

- Many different standards bodies are involved.
- There is considerable overlap of scope in this area.
- Formal (ISO) and industry standards efforts increasingly complement each other.
- ICT systems allow greater flexibility to cope with CDICT than was the case even five years ago, and standards involving Unicode, the World Wide Web, and mark-up languages are in some cases more influential than formal standards.
- Some of these features are underused.
- Many different bodies are at this present time looking at how they cope with this area.
- Most of these are tending to develop relevant standards, and also to examine their overall strategy, in isolation from other groups.

The last two of these has adversely affected the development and take up of CDICT standards, and made it more partial than would otherwise be the case, although the base standards are now more well established.

As so many groups are looking at this now, and because it is important both for European citizens, and European industry, and the global ICT industry, and because Europe provides a major part of the global market for ICT systems, and because many groups of its citizens include ICT users with similar needs to the rest of the global ICT economy, CEN/ISSS is in a good position to influence future standardization in this area.

This report will not attempt to present “definitions” neither of ***culture*** nor of ***cultural diversity***. Everyone is proud of the culture, or cultures, from which he/she grew. Culture involves language, time, place, and various related issues. A person’s culture is a vital part of the identity as an individual and as a citizen. The encouragement of cultural diversity has been a cornerstone of European relations over at least the last half-century.

It can be very easy for governments and administrations, whether at local, national, or European level, inadvertently to make a group feel threatened, and misperceptions can cause problems. Therefore enabling cultural diversity to flourish within a wider sense of European activities is vital.

Information and communication technologies (ICT) are used in every conceivable area of life. Through recent developments in standards and systems, ICT now has the potential to enable various diverse cultural characteristics to be enabled.

At the same time, these capabilities are not yet always widely understood, and there can be a risk that the way ICT systems are used can reduce cultural diversity, rather than encourage it, which can cause problems, and costs, at a wider level.

It is therefore important that ICT standards developed for all areas of European life are able to take account of the underlying capabilities enabled by recent improvements to CDICT, and that CEN enables its various standards developers to take full advantage of CDICT in their own specifications.

Europe's situation regarding cultural diversity is quite complex. Even taking together the European Union and EFTA members states, which CEN represents, this area spans three time zones, has several currencies, and many languages, using one of five European scripts (Latin, Greek, Cyrillic, Georgian and Armenian) to write them.

All these impinge on the need for standardization in Cultural Diversity in ICT systems.

Only taking into account the national languages of this area, there are around 55 languages with national or other similar official status. See annex A ("Languages in Europe").

If other areas are also included in a definition of Europe, the totals involved are even larger. For example, the Council of Europe also includes Russia and the countries of the Caucasus in Europe. The Organization for Cooperation and Security in Europe, and the Universal Postal Union also include five countries which were part of the former USSR, and which are also traditionally regarded as "Central Asia" but which have large numbers of speakers of other European languages (including Russian and to a much smaller extent, German) and which are likely to be increasingly integrated into European economies given European trade in petroleum and minerals.

The European Union has around 377 million citizens. If the EU candidate countries are added in, that totals around half a billion citizens.

There are an estimated 225 indigenous European languages, about 3 % of the world's total of around 7,000 languages. Some of those began to emerge as recognisable modern languages in Europe only in the last millennium, as peoples migrated westwards into and across Europe – as process that has been happening for at least two millennia.

Some non-indigenous languages began to be spoken in Europe in the 20th century, as some countries encouraged increased migration in order to supplement the existing workforce, and as they offered homes to refugees from across the world.

Larger cities in Western Europe have at least an additional 100–200 languages spoken as mother tongues by their school populations. The most common of these languages include Arabic, Berber, Turkish, Kurdish, Hindi, Urdu, Punjabi, and Chinese: in several European countries, these and others have a special status requiring some support for these languages in law, which in some cases may also make for additional requirements in standards relating to Cultural Diversity in ICT systems.

The Terms of Reference for the Project Team state that the Project Team should prepare a report setting out a programme of work for standardization in the area of cultural diversity and including proposals for where it should be carried out. It should be based on a consideration of:

- the recommendations of the PricewaterhouseCoopers' (PWC) report in terms of areas of activity which could benefit from standardization/consensus-building measures;
- the PricewaterhouseCoopers Annex on CEN / TC 304; and
- the CEN / TC 304 commentary.

It should also:

- recommend how the outstanding items under BC/CEN/97/26 should be progressed (if at all);
- recommend lead responsibilities for the identified work items;
- recommend future coordination of activities at the policy level including the role of industry consortia;
- recommend future coordination of activities at the technical level, including allocation and supervision of work.

Our recommendations are also made to other bodies outside of CEN/ISSS, as all those involved are part of a wider picture.

Abbreviations and other terms

CDICT – This report uses the abbreviation "CDICT" to denote that part of "cultural diversity" that is deemed to be of particular interest to the field of information and communication technology and the ICT industry.

CD – committee draft
CEN – European Committee for Standardization
DIS – draft international standard
FDIS – final draft international standard
HTML – hypertext mark-up language
ICT – information and communication technology
IEC – International Electrotechnical Commission
IS – international standard
ISO – International Organization for Standardization
JTC – joint technical committee, in particular JTC 1 between ISO and IEC
NWI – new work item
PT – project team
PWC – PricewaterhouseCoopers
PWI – preliminary work item
SC – subcommittee
SGML – Standard Generalized Markup Language
TC – technical committee
TD – Technical Direction (under ISO/IEC JTC 1)
TMX – Translation Memory eXchange format
TR – technical report
UAX – Unicode Standard Annex
UCS – Universal (Multiple-Octet Coded) Character Set
UTR – Unicode Technical Reports
UTF – Unicode (*or* UCS) Transformation Format
UTS – Unicode Technical Standard
WD – working draft
WG – working group
XLIFF – XML Localisation Interchange File Format
XML – eXtensible Markup Language

1 Strategy for Standardization of Cultural Diversity in ICT systems

The strategy should aim at improving the following:

- Formal links with standardization bodies and governments
- Interaction with the ICT industry
- Promotion to end-users of ICT and bodies serving end-users
- Relationships with the business community as ICT users
- Developing CDICT structures to reflect specific areas of Europe, among CEN member states and CEN observer states

1.1 Complexity of this domain

1.1.1 Keeping up with technology change

Standards lay down best practice in a particular area, or a means of interoperability in a particular area, based on experience of users and developers, but while the consensus process is taking place, technology moves on which may change or obsolete existing work.

An obvious example is the character set standards which CEN / TC 304 developed (and similarly many of those which ISO / TC 97 / SC 2 developed, before it became ISO/IEC JTC 1 / SC 2). These are now largely obsolete, as relatively few people are still using the technologies involved, though in some cases the continued use of legacy technology in specific areas means that the standards still need to be maintained.

Nevertheless, providing an indication of the levels of use and take-up (or otherwise) of standards relating to Cultural Diversity in ICT systems would be a useful task that a European CDICT body might undertake, as this information is very hard for users to find.

A further level of complexity is that once base standards in the CDICT area (e.g. character sets, language codes, locales, etc.) are established, other standards build on some of the basic ones. One useful task would be to assess the use of the base standards in other standards, and whether any inconsistencies are involved.

1.1.2 Involvement of National Member Bodies

Recognising areas of rapid change, and areas of low take up, compared to those of high take up, is something that all standards bodies have found it difficult to deal with, particularly those with formal National Member Body memberships, like CEN and ISO.

National Member Body memberships are very useful in dealing with strategic issues, generally representing the market players (industry, users, and administrations) at national level.

At the technical level, organizations like CEN and ISO have to assume that all committees and delegates from National Member Bodies are equally well prepared, as consensus in such a situation depends on a very formal balloting process, with outcome of results being based on a strict counting of results. However, this is not always the case.

Many industry consortia have grown up which have generally developed alternative methods of developing consensus, through attracting individual experts to prepare “standards” outside of the ISO process. This generally results in specifications being developed more quickly, but the results are often not visible to end-users, nor interoperable with other specifications.

CEN has been constructive in creating CEN Workshops, which avoids the problems of committees built around national delegations in fast-moving areas such as this. This relies on finding experts in Europe who can assess needs and develop a specification, and gain consensus among themselves in order to produce effective, stable specifications, reasonably rapidly, and CEN Workshops, and CEN Workshop Agreements (CWAs) have been a very successful development in CEN/ISSS.

1.2 Strategies on structure and cooperation

Some of the problem areas relate to where a standards body (organization, or committee, or other) fits into the general scheme of standardization. In most cases this is because the body has become somehow disconnected from its user groups, or from other "natural allies" who would cooperate with it in developing standards. One example in ISO is in the bibliographic character set standards developed originally by ISO / TC 46 / SC 4 / WG 1, which have had marginal take-up, and where the industry has always used a completely different de facto standard, regardless of the technical activity in ISO.

There are however, some areas of success in this area, in some ISO committees. In passing, some of these areas of success have been when ISO and consortia, or in some cases ISO and intergovernmental bodies, have deliberately cooperated.

Indeed, formalizing cooperation between CEN/ISSS and selected other bodies in this domain, and at the right level(s), could well be one of the significant recommendations.

1.3 TC/Consortia cooperation

Examples of successful TC/Consortia cooperation include are

(a) ISO/IEC JTC 1 / SC 2 / WG 2 and the Unicode Consortium, in developing UCS (ISO/IEC 10646 and Unicode),

(b) ISO / TC 37 and LISA, in developing various XML/DTDs as standards in specific application areas, in machine translation, localization, and terminology, and

(c) CEN / TC 304 and the Unicode Consortium.

1.4 TC/Intergovernmental Body cooperation

Examples of TC/Intergovernmental Body cooperation include

(a) ISO/IEC JTC 1 / SC 17 / WG 3 in developing ISO/IEC 7501 (*Machine readable travel documents*) together with the International Civil Aviation Organization's Technical Advisory Group (ICAO TAG), and

(b) ISO/IEC JTC 1 / SC 32 together with UN/CEFACT in Geneva, in developing various standards related to EDI.

CEN/ISSS activities do involve some collaboration with consortia, and this is generally increasing. In addition, we have close links with the European Union and EFTA and sometimes with Governmental bodies – for instance, CEN/ISSS WS/eBES is the European Entry Point to the UN/CEFACT system. CEN/ISSS is therefore potentially in a strong position to influence events, and various consortia meetings have sometimes bemoaned the lack of any input from CEN/ISSS at some of their meetings.

2 Review of the current situation

This chapter gives an overview of the current situation in the field of cultural diversity and ICT. We have in particular looked at standardization activities both inside and outside the European and international standardization organizations.

The three most important areas of standardization in the areas of CDICT are Unicode, the World Wide Web (WWW), and various mark-up languages, particularly XML these days, as well as base standards (mainly in ISO/IEC JTC 1 SCs) relating to these. These affect each other, and also affect many other developments in ICT, including CDICT.

Much of the information below has been collected from written sources, chiefly on the Internet. A brief questionnaire was circulated by email to secretaries, chairmen and other contacts. Some of the analysis is based on feedback from this questionnaire. A number of experts have been contacted directly. The Project Team is grateful for all information that we have received. Regretfully, some experts have not responded. We hope that we have been able to give an accurate description of "their" groups and activities in spite of this.

Some organizations that were intended for inclusion in this overview have been taken out from the final version of this report. Some of the organizations have a rather peripheral engagement in the field of CDICT; other organizations have unfortunately not provided the relevant information to the Project Team.

Overviews of (parts of) CDICT are available through links under many of the web pages that are referenced here. One particularly interesting source of such overview is the Diffuse web site – http://www.diffuse.org/ – hosted by the European Commission.

2.1 CEN and CENELEC

Within a European context, Cultural Diversity in ICT standardization in CEN has centred on CEN / TC 304, which grew out of a character set committee – much as did ISO/IEC JTC 1 / SC 2 / WG 2, which has had to cover more aspects due to the interrelationship of ISO/IEC 10646 and Unicode. In both cases, a traditional model of consensus among delegates from National Member Bodies has been used. Consensus was not always available in relation to some standards.

CEN / TC 304 has also set up Project Teams to develop specific standards and technical reports, the results to be approved by CEN / TC 304 and its National Member Bodies, and also set up some Workshops.

In more recent years, CEN/ISSS has set up Workshops for limited times, involving individual experts, and addressed to specific tasks. Some of these have lasted longer, covering more ground, such as the eCommerce, eLearning and Dublin Core/Metadata Workshops.

In passing, there are also some CENELEC committees which have covered data elements used in electrical standards, which may have some relevance to this study, though these committees have functioned more as users of base standards than as initiating development.

Most CEN / TC 304 standards have related in one way or another to ISO/IEC 10646, the repertoire of which is synchronized with Unicode.

2.1.1 Technical Committee(s)

2.1.1.1 CEN / TC 304 – *Information and communications technologies – European localization requirements*

Starting with a focus on character sets, CEN / TC 304 developed into a "CDICT committee". There are very few parts of CDICT that fall outside the area of interest of TC 304.

Scope

Standardization in the field of Information and Communications Technologies, to ensure that European localization requirements can be satisfied. Localization in this context means the provision of software and hardware support adapted to local linguistic and cultural needs in Europe.

Structure

CEN / TC 304 currently has a very low activity level. A few ongoing projects are being finalized under an interim secretariat. It has been decided to discontinue the committee following the finalization of some ongoing projects.

CEN / TC 304 was subdivided into Working Groups to begin with. After a re-organization it had no formal subdivision. Small Project Teams were established to draft documents. The European Commission funded some of the activities.

It needs to be noted that this structure (after the discontinuation of the Working Groups) made it somewhat difficult to find a good place for some of the technical discussions. Formal issues were discussed in the TC meetings, while technical issues were frequently discussed in closed Project Team meetings only.

Projects with relevance to CDICT

Most or all projects under CEN / TC 304 has a direct relevance to CDICT. These projects are well known to CEN/ISSS, and we give just a brief list:

- character sets
- keyboards
- matching, browsing
- alphabetical ordering
- requirements for linguistic groups (e.g. Sami)

See Annex B for a list of deliverables from CEN / TC 304.

2.1.1.2 WS/Alpha – Alphabets of European Languages Workshop

Scope

This CEN/ISSS Workshop was created to deliver a catalogue of the alphabets used to represent the indigenous languages of Europe, which has been reviewed according to a set of criteria also to be established by the Workshop. The catalogue was intended to be used by standardization bodies and IT-industry developers as a source for the alphabets used by the indigenous languages of Europe. The resulting CWA was agreed to be the agreed upon criteria and review process and a list of non-controversial European Alphabets and letters.

The Workshop was announced to be closed down following the impossibility to reach consensus on the CWA, given that there were two quite diametrically opposed views: a majority that was in favour of the inclusion of some 160+ languages (and dialects others would say) and a minority (but sufficiently important) that wanted to restrict the languages to those referenced in the European Charter for Regional and Minority Languages.

2.1.1.3 WS/ESR – European Culturally Specific Requirements Workshop

Scope

This CEN Workshop defines a check list of Culturally Specific ICT Requirements, such as character sets, internationalisation and user interfaces, in Europe, that products and services developed on the framework of the Global Information Infrastructure need to cover and support.

The CWA also discusses the rationale for the requirements that affect the localisation of ICT systems and services. In addition to the requirements in a national / cultural application environment, the CWA identifies areas where national requirements still need be addressed even in pan-European applications.

Structure

The CEN/ISSS ESR Workshop started its activities in November 1998 and was closed end 2000, following the completion of its work, with the approval of a CWA.

Projects with relevance to CDICT

CWA 14094 European Culturally Specific ICT Requirements

2.1.1.4 WS/Eurolocale

Scope

The Workshop defines the generic language-neutral locale for Europe, in the form of a Narrative Cultural Specification, plus generic language-dependent locales for each of the official and treaty languages of the European Union and EFTA. It is intended for use in European institutions, and also as a base for modification for national specifications, that only deviate relatively little. The components of the CWA are intended for registration with the cultural registration standard ISO/IEC 15897.

Structure

The CEN/ISSS Eurolocale Workshop started its activities in June 1998 and was closed end 2000, following the completion of its work, with the approval of a CWA.

Projects with relevance to CDICT

The CWA consists of two parts:

CWA 14051-1 – *Information Technology – European generic locales Part 1: General specifications.*

CWA 14051-2 – *Information Technology – European generic locales Part 2: Narrative cultural specifications, POSIX locales, and repertoiremap.*

2.1.1.5 WS/MES – *Multilingual European Subsets*

Scope

The Workshop defines European subset repertoires on the basis of ISO/IEC 10646.

Structure

The CEN/ISSS MES Workshop started its activities in June 1998 and completed its work in December 1999, with the approval of a CWA.

Projects with relevance to CDICT

The MES CWA has been published as CWA 13873:2000. In this CWA, the following subsets are specified:

MES-1: A Latin repertoire based on ISO/IEC 6937:1994 (a limited subset, fixed collection).

MES-2: A Latin, Greek, Cyrillic repertoire based on ENV 1973:1996 (a limited subset, fixed collection).

MES-3: A repertoire needed to write the languages of Europe and transliterate between them (a selected subset (non-fixed collection – referred to as **MES-3A**) and a limited subset (fixed collection – referred to as **MES-3B**))

2.1.1.6 WS/EC – *Electronic commerce*

Web site: http://www.cenorm.be/isss/Workshop/ec/.

This generic Workshop has been established to provide an open and flexible framework for market players (manufacturers, service providers, users, research bodies, administrations, etc.) to identify and progress eCommerce standards and standards-related issues, and to deliver outputs. As eCommerce impacts on all business processes, a common, multi-sectoral approach is crucial for resolving interoperability between technical solutions, which need to be implemented across businesses and value chains. The Workshop offers a coherent and cohesive focus for EC standardization at a European level, within the context of global EC standardization activities.

The following CEN Workshop Agreements have been published:

- **CWA 14228** – *Summaries of some Frameworks, Architectures and Models for Electronic Commerce.*
- **CWA 14162** – *Datatyping for Electronic Data Interchange.*
- **CWA 13993** – *Recommendations and Guidance on the use of XML for Electronic Data Interchange.*
- **CWA 13992** – *Recommendations for Standardization in the field of XML Electronic Data Interchange.*
- **CWA 13692** – *PBDH (Product and Business Data Administration): A common object repository of EC/EDI types.*
- **CWA 13691** – *MIGs (Message Implementation Guidelines): A web-based repository which indexes or holds MIGs.*

2.1.1.7 WS/LT – *Learning technologies*

Web Site: http://www.cenorm.be/isss/Workshop/LT/.

A number of, if not all of, the LT Workshop work items have direct CDICT impact:

- Internationalisation of the IEEE LTSC Learning Object Metadata (LOM) specification
- Availability of alternative language versions of a learning resource in the IEEE LTSC Learning Object Metadata (LOM) specification
- Translation of LOM into various European Languages
- Description of language capabilities
- Quality assurance
- Repository of taxonomies/vocabularies for a European Learning Society
- Educational Copyright Licence Conditions
- Educational modelling languages
- Description of Learning Resource Capabilities with respect to Accessibility Requirements
- Interoperability frameworks for exchange of information between diverse management systems
- Handling of Learner Profiles in IT-supported learning environments from a European perspective
- Learning Technology Standards Observatory
- Knowledge Content Interoperability Framework

The following CEN Workshop Agreement has been published so far:

- **CWA 14040** – *A Standardization Work Programme for "Learning and Training Technologies & Educational Multimedia Software".*

2.1.1.8 WS/KM – *Knowledge management*

Web site: http://www.cenorm.be/isss/Workshop/km/.

The Workshop on Knowledge Management has been established to develop a "European guide to Good Practice in Knowledge Management". The Workshop was formally launched at the Kick-Off meeting on 2002-06-24 in Brussels. A Project Team is currently being established.

2.1.1.9 WS/eBES – *e-Business Board for European Standardization*

Web site: http://www.cenorm.be/isss/Workshop/eBES/.

The former EBES Workshop was merged with the XML/EDI Workshop in 2001 to form a new eBES Workshop. The major objective of the new eBES Workshop is to create within Europe a central point focusing on the latest technologies used for the exchange of electronic business data.

eBES will represent the European point of view in the global debate, including on the need for a multi-lingual and multi-cultural approach to B2B data interchange standardization. This central focal point will provide information about new standardized technologies in this field, will foster their use, will participate in or provide input to the global processes defined by ebXML, OASIS and UN/CEFACT. Particular emphasis is placed on the role SMEs can play, without neglecting the large enterprises who already made huge investments in eCommerce and want to exploit these further.

2.1.1.10 WS/MMI-DC – *MMI - Dublin Core*

Web site: http://www.cenorm.be/isss/Workshop/MMI-DC/.

The Dublin Core has become an important part of the emerging infrastructure of the Internet. The MMI-DC Workshop has established itself as a recognized player in relation to metadata from a Dublin Core perspective and plays a visible role in the promotion of a Dublin Core based metadata set as a common core to applications that use metadata.

The following CEN Workshop Agreements have been delivered so far:

- **CWA 13989** – *Description of structure and maintenance of the web based Observatory of European work on metadata.*
- **CWA 13988** – *Guidance on use of Dublin Core in Europe.*
- **CWA 13874** – *endorsing Dublin Core Metadata Element Set Version 1.1.*

2.2 ISO/IEC JTC 1 – *Information technology*

Central web site: http://www.jtc1.org/. Each of the committees has their own web sites, which are linked from the central web site.

Scope

The overall scope of ISO/IEC JTC 1 is simply: *Standardization in the field of information technology*.

Structure

ISO/IEC JTC 1 is divided into 17 Subcommittees, each with numerous Working Groups. In many respects JTC 1 Subcommittees are on the level of ISO Technical Committees.

JTC 1 SCs are grouped into 12 Technical Directions (TD), most of which has one or two SCs. One of the SCs is split between two TDs: SC 22 (*Programming languages, their environments and systems software interfaces*) is in TD *"Programming languages and software interfaces"*, while its WG 20 (*Internationalization*) is in TD *"Cultural and linguistic adaptability and user interfaces"*.

The structuring element "Technical Direction" is not found in other parts of the ISO/IEC system. It seems to be a result of an attempt to group the activities into thematic groups. However, since most of the TDs have just one or two SCs, the effect of the structuring into TDs isn't obvious.

The TD "CLAUI" is of particular interest for CDICT. The following sub-groups belong to this TD: SC 2, SC 22 / WG 20, and SC 35. These and other groups are described below.

In ISO/IEC JTC 1, standardization efforts for Cultural Diversity in ICT have been spread among several committees: ISO/IEC JTC 1 / SC 2 / WG 2, ISO/IEC JTC 1 / SC 22 / WG 20 and ISO/IEC JTC 1 / SC 35.

Groups under other TDs as well have relevance. The following are included in the description in this report: Application Technologies: SC 36; Data Capture and Identification Systems: SC 17; Data Management Services: SC 32.

2.2.1 CLAUI – *Cultural and linguistic adaptability and user interfaces*

Some CLAUI documents are available at http://std.dkuug.dk/jtc1/claui/.

The Cultural and Linguistic Adaptability and User Interface Technical Direction (CLAUI TD) was set up by ISO/IEC JTC 1 to oversee this. It is to be noted that constructive collaboration with the Unicode Consortium has been successful in some standards, particularly in two standards committees, respectively in JTC 1 / SC 2 and JTC 1 / SC 22 / WG 20. This is explainable because the two involved standards relate to the very infrastructure by which we will all communicate in the future, that of the universal character set (ISO/IEC 10646), closely related to several aspects of Unicode, and identical in its coding, and its not less essential complement, the ordering of the characters and scripts of the world (ISO/IEC 14651), a sensitive processing aspect in which tailoring plays an important and mandatory role (in this regard, the Unicode Collation Algorithm [UCA] and its default collation table constitute an important practical delta (profile) of ISO/IEC 14651; its ordering table is

identical to the template table of ISO/IEC 14651, a table obtained by international consensus in ISO/IEC JTC 1, then fed back into "Unicode Unisort tables", the ISO/IEC 14651 table itself currently being then generated by an automatic process out of Unicode data, a process that is done very quickly after international ballots; this is an example of synchronized work).

Up to now, CDICT has not formally featured in the scope of ISO/IEC JTC 1, but early indications from a panel looking at the future of JTC 1 structures indicate that the scope statement for ISO/IEC JTC 1 may now include reference to "related cultural and linguistic adaptability and societal aspects" as well as to other aspects of ICT standardization.

CLAUI TD was an earlier attempt to bring together CDICT standardization activities within ISO/IEC JTC 1. However, rather than bringing them together in a single SC, ISO/IEC JTC 1 decided instead to bring them together in a TD although this is not a method of working that is well-defined in ISO/IEC JTC 1.

In addition, CLAUI TD coordination meetings have not always been well attended (notably, ISO/IEC JTC 1 / SC 2 has sporadically been absent), with some difficulties to produce a consistent strategic approach.

ISO/IEC JTC 1 / SC 2, JTC 1 / SC 22 / WG 20, and JTC 1 / SC 35 are the constituent committees of the CLAUI TD.

ISO/IEC JTC 1 / SC 35 and ISO/IEC JTC1/SC22/WG20 have been active constantly in CLAUI TD, whereas ISO/IEC JTC 1 / SC 2 / WG 2 has felt less concerned for different reasons

ISO/IEC JTC 1 / SC 2 / WG 2 and ISO/IEC JTC 1 / SC 22 / WG 20 have had common technical involvement in one important standard, ISO/IEC 14651 on ordering (an involvement which has worked remarkably well along with Unicode Technical Committee's commitment over the last 5 years, after some previous years of important struggle between the tow bodies).

2.2.2 ISO/IEC JTC 1 / SC 2 – *Coded character sets*

Web site of SC 2: http://std.dkuug.dk/jtc1/sc2/.

Area of work

Standardization of graphic character sets and their characteristics, associated control functions, their coded representation for information interchange, and code extension techniques.

Structure

There are two active Working Groups:

- WG 2 – *Multi-octet codes*
- WG 3 – *7 and 8 bit codes and their extension*

WG 3 is relatively inactive, although a number of its International Standards are still highly relevant. WG 2 is the home of ISO/IEC 10646 (cf Unicode).

Projects with relevance to CDICT

- **ISO/IEC 10646-1:2000** – *Information technology – Universal Multiple-Octet Coded Character Set (UCS) – Part 1: Architecture and Basic Multilingual Plane.*
- **ISO/IEC 10646-2:2001** – *Information technology – Universal Multiple-Octet Coded Character Set (UCS) – Part 2: Supplementary Planes.*

All parts of ISO/IEC 10646 are aligned with Unicode, and the repertoire is synchronized. The Unicode Standard standardizes other aspects besides just character coding. See clause 2.8 on Unicode for an overview of both.

Most significant development is done on ISO/IEC 10646 within ISO/IEC JTC 1 / SC 2 / WG 2.

Most of the work in ISO/IEC JTC 1 / SC 2 / WG 3 is on legacy character sets:

- 8-bit coded character sets which provide 8-bit sub-repertoires of ISO/IEC 10646 (principally all parts of ISO/IEC 8859, each part providing the repertoire for one or more languages, in various scripts);
- 7-bit and 8-bit character set mechanisms;

- 7-bit and 8-bit bibliographic character sets taken over from ISO / TC 46 / SC 4, most of which are little used in practice, even in libraries, and which are not being developed further.

In a sense, the following can be regarded as legacy sets, in that they are not being developed further: on the other hand they are extremely current standards, due to their widespread use, even if it is unlikely to change.

- Control character sets: most controls in use are now well established and not much added too, and can be used with any character set. Control codes and control sequences specified in ISO 6429 also allow for bidirectional text.

These are used principally in machine-readable passports and similar machine-readable identity cards. Currently only 7-bit ("ASCII") characters from ISO/IEC 646 are in use. CEN / TC 304 has promoted the development the repertoire to allow for accented characters.

However, machine-readable passports contain both a machine-readable zone (which only requires to be read by an OCR reader) and a human readable zone, which is produced using physical writing or printing techniques and which can contain information in any language or script. Name information does not need to be identical in both zones.

For further information, see also clause 2.2.4 on JTC 1 / SC 17 / WG 3 (on machine-readable passports) and clause 2.3.3 on ISO / TC 46 / WG 3 (previously ISO / TC 46 / SC 2, Conversion of written languages) on transliteration of names.

See Annex C for a list of relevant International Standards from ISO/IEC JTC 1 / SC 2.

2.2.3 ISO/IEC JTC 1 / SC 17 – *Cards and personal identification*

The SC has its own web site: http://www.sc17.com/.

Scope

Standardization in the area of (a) identification and related documents, (b) cards, and devices associated with their use in interindustry applications and international interchange.

Structure

ISO/IEC JTC 1 / SC 17 has eight Working Groups:

- WG 1 – *Physical characteristics and test methods for ID-cards*
- WG 3 – *Identification cards – Machine readable travel documents*
- WG 4 – *Integrated circuit card with contacts*
- WG 5 – *Registration Management Group (RMG)*
- WG 8 – *Integrated circuit cards without contacts*
- WG 9 – *Optical memory cards and devices*
- WG 10 – *Motor vehicle driver licence and related documents*
- WG 11 – *Biometrics* (newly established; its first meeting scheduled for 2002-06-26/27)

Projects with relevance to CDICT

Only the standards issued by ISO/IEC JTC 1 / SC 17 / WG 3 are relevant to CDICT. These are the three parts of ISO/IEC 7501:

- **ISO/IEC 7501-1:1997** – *Identification cards – Machine readable travel documents – Part 1: Machine readable passport.*
- **ISO/IEC 7501-2:1997** – *Identification cards – Machine readable travel documents – Part 2: Machine readable visa.*
- **ISO/IEC 7501-3:1997** – *Identification cards – Machine readable travel documents – Part 3: Machine readable official travel documents.*

Development of all parts of ISO/IEC 7501 is controlled by the International Civil Aviation Organization in Montreal, through its Technical Advisory Group (ICAO TAG). A few members of ISO/IEC JTC 1 / SC 17 / WG 3 are permitted to attend as observers on ICAO TAG, and which only government representatives are able to serve fully.

ISO/IEC 7501 is tied to the revision of the ICAO standard numbered as ICAO Doc. 9303, which it appears national governments have a legal obligation to implement under international civil aviation law.

A guide to that document can be found at http://www.icao.int/icao/en/atb/fal/mrtd/guide.htm.

ICAO Doc. 9303 determines ISO/IEC 7501 rather than the other way around. It provides specifications for Machine Readable Travel Documents, and is similarly published in three parts:

- Part 1 Machine Readable Passports
- Part 2 Machine Readable Visas
- Part 3 Machine Readable Official Travel Documents

This includes tables for transliterations to be used in passports, which is proving controversial for some national governments as well as users. ISO/IEC JTC 1 / SC 17 / WG 3 set up a Task Force on Transliteration, and decided to produce tables independently. Normal liaison relation between the two WGs did not produce the desired results in this case. Similar lack of cooperation is not uncommon between ISO committees and other committees that have relevant expertise.

Transliteration in passports is an area where CDICT can have legal implications: court cases in Greece, Germany and Latvia in recent years have all turned on representation of personal names in machine-readable and/or official documents.

A description of CDICT issues in this area can be found in the legal decisions of the Constitutional Court of Latvia, at http://www.satv.tiesa.gov.lv/Eng/Spriedumi/04-0103(01).htm.

There is also a need for government agencies to be able to track variants of personal names in databases, to avoid legal loopholes being exploited in travel or various official dealings.

It may be expected that projects stemming from the new WG 11 will also be highly relevant to CDICT.

2.2.4 ISO/IEC JTC 1 / SC 22 / WG 20 – *Internationalization*

SC 22 / WG 20 was formed within ISO/IEC JTC 1 / SC 22 (*Programming languages*) to address common issues of internationalization (i18n) in programming languages and system interfaces, including character sets, cultural conventions, and culturally correct ordering. It must be remembered that the WG title "means" Internationalization in the context of the overall scope of SC 22, which is also reflected in the Area of work.

The WG has its own web site: http://std.dkuug.dk/JTC1/SC22/WG20/.

Area of work

Identification of elements relevant to the work of SC22 (programming languages, their environments and system software interfaces) that may be affected by differences in language, culture, customs and habits; and for these elements, develop standards that enable applications to be portable across differing cultural practices; and develop a Technical Report that describes a framework for nations to provide those elements.

Structure

ISO/IEC JTC 1 / SC 22 / WG 20 is not subdivided.

Projects with relevance to CDICT

Most WG20 standards are relevant to CDICT. Published standards include:

- **ISO/IEC TR 10176:2001** – *Information technology – Guidelines for the preparation of programming language standards.*
- **ISO/IEC TR 11017:1998** – *Information technology – Framework for internationalization.*

These two technical reports describe user needs and developer needs in general terms.

Annex A of ISO/IEC TR 10176 needs constant updates due to the expanding repertoire of ISO/IEC 10646, in Annex A (*Characters to be used in identifiers*) of ISO/IEC TR 10176.

- **ISO/IEC 14651:2001** – *Information technology – International string ordering and comparison – Method for comparing character strings and description of the common template tailorable ordering.* – This is a

basic standard for the description of alphabetical ordering procedures. The project needs constant updates due to the expanding repertoire of the Universal Character Set (UCS, ISO/IEC 10646). Updating, in step with the Unicode Collation Algorithm, comprises the largest part of the work of JTC 1 / SC 22 / WG 20. ISO/IEC 14651 is implemented in the European Ordering Rules (see under CEN / TC 304) and in ISO 12199 (see under ISO / TC 37). Tailoring (albeit expected to be very light in a given local environment) constitutes a mandatory process to conform to ISO/IEC 14651, an important feature serving the cultural diversity preservation. All main ordering schemes are about to be built as deltas (profiles) of ISO/IEC 14651 (including the UCA, which is fed back in both directions, from and into ISO/IEC, after consensus ballots).

- **ISO/IEC TR 14652** – *Specification method for cultural conventions.* – This failed standard, which nevertheless brings important notions for the future of internationalisation support, has finally been approved at the ISO/IEC JTC 1 level and will be published.
- **ISO/IEC 15897:1999** – *Information technology – Procedures for registration of cultural elements.* – This project was fast tracked from ENV 12005 and assigned to WG 20 for maintenance. A major revision is in preparation. The registration authority is DKUUG (Denmark).

2.2.5 ISO/IEC JTC 1 / SC 32 – *Data management and interchange*

The SC has its web site at http://www.jtc1sc32.org/.

Area of work

Standards for data management within and among local and distributed information systems environments. SC 32 provides enabling technologies to promote harmonization of data management facilities across sector-specific areas.

Specifically, SC 32 standards include:
1) reference models and frameworks for the coordination of existing and emerging standards;
2) definition of data domains, data types and data structures, and their associated semantics;
3) languages, services and protocols for persistent storage, concurrent access, concurrent update and interchange of data;
4) methods, languages, services and protocols to structure, organize and register metadata and other information resources associated with sharing and interoperability, including electronic commerce.

Structure

ISO/IEC JTC 1 / SC 32 has five Working Groups and one “RG” (for the maintenance of ISO/IEC 10032).

- WG 1 – *Open EDI*
- WG 2 – *MetaData*
- WG 3 – *Database language*
- WG 4 – *SQL/Multimedia and application packages*
- WG 5 – *Database access and interchange*

Projects with relevance to CDICT

In particular metadata projects are highly relevant, including the **ISO/IEC 11179** series, the **ISO/IEC 13249** series, and the **ISO/IEC 15944** series. See Annex D for a list of ISO/IEC JTC 1 / SC 32 standards that may be relevant.

2.2.6 ISO/IEC JTC 1 / SC 35 – *User interfaces*

The SC 35 web site can be accessed from http://www.jtc1.org/ and directly at http://forum.afnor.fr/afnor/WORK/AFNOR/GPN2/Z62A/index.htm.

Scope

Standardization in the field of interfaces between users (including people with special needs) and system, encompassing input and output devices in information technology environments, with the priority of meeting the JTC 1 requirements for cultural and linguistic adaptability.

The standardization work on SC 35 includes the following areas:

- *Interfaces between users and devices, such as keyboards, mice, pointers, pens, visual displays, and forms of audio and tactile input/output with the emphasis on functionality.*
- *Rules for system control by voice, vision, movement, gestures, etc.*
- *Presentation techniques, graphical symbols, icons etc.*
- *Dialogue control and navigation in interactions between humans and systems assistance and tutoring.*

Structure

- WG 1 – *Keyboards and input interfaces*
- WG 2 – *User interface interaction*
- WG 3 – *Graphical symbols*
- WG 4 – *User interfaces for mobile devices*
- WG 5 – *Cultural, linguistic and user requirements*
- WG 6 – *User interfaces for disabled and elderly people*

Projects with relevance to CDICT

Most or all projects under ISO/IEC JTC 1 / SC 35 are relevant from a CDICT point of view.

Its most known series of standards, ISO/IEC 9995, defines important aspects in keyboard design, guidelines for the design of national keyboards, and goes as far as being the international reference for allocation Latin letters to the keys of numeric keypads, whose most notorious applications are in use on telephone sets and automatic bank teller machines.

ISO/IEC JTC 1 / SC 35 has recently created two promising working groups, WG 5 focusing directly on establishing recommended practices on how to evaluate cultural and linguistic adaptability in ITC products (project TR 19764 on “Guidelines, methodology, and reference criteria for cultural and linguistic adaptability in information technology products”, a project extremely relevant as a tool to ensure cultural diversity in the future), and WG 6 focusing on “User interfaces for people with special needs (including the elderly and disabled)”. Its other WGs work on icons, symbols, mobile personal assistants, and the general user-system interface, and also produce standards with a positive impact on cultural and linguistic adaptability.

See Annex E for a complete list of International Standards produced by ISO/IEC JTC 1 / SC 35.

2.2.7 ISO/IEC JTC 1 / SC 36 – *Learning technology*

This SC has been formed quite recently. It had its first meeting in 2000. The SC has its web site at http://jtc1sc36.org/.

Area of work

Standardization in the area of information technologies that support automation for learners, learning institutions, and learning resources. Excluded: The SC shall not create standards or technical reports that define educational standards, cultural conventions, learning objectives, or specific learning content.

Structure

ISO/IEC JTC 1 / SC 36 has three working groups. In addition, ad hoc groups are formed as needed.

- WG 1 – *Vocabulary*
- WG 2 – *Collaborative technology*
- WG 3 – *Learner information*

Projects with relevance to CDICT

- **ISO/IEC CD1 21484-13** – *Information technology – Learning, education, and training – Simple human identifiers.*

2.3 ISO committees

2.3.1 ISO / TC 12 – *Quantities, units, symbols, conversion factors*

This technical committee is included because of its fundamental and cross-sectoral nature. No particular reference is made to the committee in other parts of this report.

Scope

Standardization of units and symbols for quantities and units (and mathematical symbols) used within the different fields of science and technology, giving, where necessary, definitions of these quantities and units. Standard conversion factors between the various units.

Structure

ISO / TC 12 has no subcommittees. It has four active working groups.

Projects with relevance to CDICT

All standards from ISO / TC 12 are relevant to all areas of standardization.

2.3.2 ISO / TC 37 – *Terminology and other language resources*

Home page maintained by ISO / TC 37 Secretariat: http://linux.infoterm.org/iso-e/i-iso.htm.

The title and scope of ISO / TC 37 was changed in 2000. The previous scope included terminology, terminography, and lexicography, while the new scope is extended to include all language resources.

Current scope of ISO / TC 37

Standardization of principles, methods and applications relating to terminology and other language resources.

Scope of subcommittees

SC 1 – Principles and methods: *Standardization of basic principles and methods for developing scientific and technical terminologies and other language resources.*

SC 2 – Terminography and lexicography: *Standardization of the application of principles and methods in terminology work, terminography and lexicography.*

SC 3 – Computer applications: *Standardization of models for information processing and of related coding systems applicable to terminology work and terminography.*

SC 4 – Language resource management: *Standardization of specifications for computer-assisted language resource management.*

Structure

ISO / TC 37 is subdivided into four Subcommittees. SC 1, 2, and 3 have 3–4 Working Groups. SC 4 is newly established, and is not (yet) subdivided. There are no WGs directly under the TC.

The TC and the SCs have permanent chairmen and secretariats. All active WGs have permanent conveners. Each active project has a project leader. The conveners are normally also project leaders for one or more projects under the WG.

The TC and its sub-groups usually has one meeting per year. The level of activity between meetings depends largely on the project leaders.

Projects with relevance to CDICT

Principles and methods within terminology and language resources are highly relevant to CDICT, although some of the projects have a more indirect relevance.

- **PWI** – *Basic principles of multilingual product classification for electronic commerce.* – This project will focus on multilingual and multicultural aspects of electronic commerce.
- **ISO 639** – *Codes for the representation of names of languages.* – The tables (alpha-2 and alpha-3 codes) are being continuously updated; see the home page of the Joint Advisory Committee: http://lcweb.loc.gov/standards/iso639-2/iso639jac.html. Latest publications: **ISO 639-1:2002** – … *Part 1: Alpha-2 code,* and **ISO 639-2:1998** – … *Part 2: Alpha-3 code.*
- **ISO 704:2000** – *Terminology work – Principles and methods.* – This and the following two are among the basic standards for terminology work.
- **ISO 860:1996** – *Terminology work – Harmonization of concepts and terms.*
- **ISO 1087** – *Terminology work – Vocabulary.* – Published in two parts: **ISO 1087-1:2000** – … *Part 1: Theory and application,* and **ISO 1087-2:2000** – … *Part 2: Computer applications.*
- **ISO 1951:1997** – *Lexicographical symbols and typographical conventions for use in terminography.*
- **ISO 10241:1992** – *International terminology standards – Preparation and layout.*
- **ISO 12199:2000** – *Alphabetical ordering of multilingual terminological and lexicographical data represented in the Latin alphabet.*
- **ISO 12200:1999** – *Terminology – Computer applications – Machine-readable terminology interchange format (MARTIF) – Negotiated interchange.*
- **ISO 12616:2002** – *Translation-oriented terminography.*
- **ISO/TR 12618:1994** – *Computational aids in terminology – Creation and use of terminological databases and text corpora.*
- **ISO 12620:1999** – *Terminology – Computer applications – Data categories.* – Basic "metadata" standard, forming together with 12200, 16642, 16503, and 17241 generic format descriptions for terminology.
- **ISO/DIS 16642** – *Computer applications in terminology – Terminological mark-up framework (TMF).*
- **ISO/DIS 16503** – *Computer applications in terminology – Representation format for terminological data collections – MARTIF-compatible with specified constraints (MSC).*
- **ISO/DIS 17241** – *Computer applications in terminology – Generic model (GENETER) for SGML-based representation of terminological data.*

2.3.3 ISO / TC 46 – *Information and documentation*

Activities in this committee are highly relevant to the scope of this study. Regretfully, officers in the committee have not responded to our enquiries. The information below has been derived from written sources (in particular on the ISO web site) and from PT members' knowledge about the committee. There is no Business Plan for public review available for ISO / TC 46.

Scope

Standardization of practices relating to libraries, documentation and information centres, indexing and abstracting services, archives, information science and publishing.

Structure

ISO / TC 46 has the following working groups and subcommittees:

- WG 2 – Coding of country names and related entities
- WG 3 – Conversion of written languages
- WG 4 – Terminology of information and documentation
- SC 4 – Computer applications in information and documentation
- SC 4 / WG 1 – Character sets
- SC 4 / WG 4 – Format structures for bibliographic information interchange in machine readable form
- SC 4 / WG 6 – Electronic publishing

- SC 4 / WG 7 – Data elements
- SC 4 / WG 8 – Library codes
- SC 8 – Quality - Statistics and performance evaluation
- SC 8 / WG 2 – International library statistics
- SC 8 / WG 4 – Performance indicators for libraries
- SC 9 – Identification and description
- SC 9 / WG 1 – International standard audiovisual number (ISAN)
- SC 9 / WG 2 – International standard work code (ISWC)
- SC 9 / WG 3 – International standard textual work code (ISTC)
- SC 9 / WG 4 – International standard book numbering (ISBN)
- SC 11 – Archives/records management

Activities within most of these groups have relevance to CDICT, in particular: WG 2, WG 3, SC 4 / WG 1, and SC 4 / WG 7.

Projects with relevance to CDICT

All projects relating to ***character sets and transliteration*** are relevant: **ISO 9:1995** (Cyrillic transliteration); **ISO 233:1984** (Arabic transliteration); **ISO 233-2:1993** (Arabic transliteration); **ISO 233-3:1999** (Arabic/Persian transliteration); **ISO/DIS 233-4** (Arabic/Persian transliteration); **ISO 259:1984** (Hebrew transliteration); **ISO 259-2:1994** (Hebrew transliteration); **ISO 843:1997** (Greek transliteration); **ISO 3602:1989** (Japanese kana romanization); **ISO 5426-2:1996** (extension of the Latin alphabet); **ISO 6861:1996** (Glagolitic alphabet coded character set); **ISO 7098:1991** (Chinese romanization); **ISO 8957:1996** (Hebrew alphabet coded character sets); **ISO 9984:1996** (Georgian transliteration); **ISO 9985:1996** (Armenian transliteration); **ISO 11940:1998** (Thai transliteration); **ISO/DIS 11940-2** (Thai transliteration); **ISO/TR 11941:1996** (Korean romanization); **ISO 10754:1996** (extension of the Cyrillic alphabet coded character set); **ISO 15919:2001** (transliteration of Devanagari and related Indic scripts).

Other projects include:

- **ISO 3166-1:1997** – *Codes for the representation of names of countries and their subdivisions – Part 1: Country codes.*
- **ISO 3166-2:1998** – *Codes for the representation of names of countries and their subdivisions – Part 2: Country subdivision code.*
- **ISO 3166-3:1999** – *Codes for the representation of names of countries and their subdivisions – Part 3: Code for formerly used names of countries.* – All three parts of this standard are relevant. It is being continuously maintained; see Maintenance Agency home page: http://www.iso.org/iso/en/prods-services/iso3166ma/.
- **ISO 5963:1985** – *Documentation – Methods for examining documents, determining their subjects, and selecting indexing terms.* – Relevant for multilingual indexing and search.
- **ISO/DIS 15924** – *Information and documentation – Code for the representation of names of scripts.* – This is a central document that is closely related to ISO 639 (see under ISO / TC 37).
- **ISO/AWI 23081** – Metadata for records and records management processes.

See Annex F for a list of ISO / TC 46 standards.

2.3.4 ISO / TC 59 – *Building construction*

This committee is included as an example of a committee that focuses on technical aspects that nevertheless may have CDICT implications.

Scope

Standardization in the field of building and civil engineering, of:

- *general terminology for building and civil engineering;*
- *organization of information in the processes of design, manufacture and construction;*
- *general geometric requirements for building, building elements and components including modular coordination and its basic principles, general rules for joints, tolerances and fits;*
- *general rules for other performance requirements for buildings and building elements including the coordination of these with performance requirements of building components to be used in building and civil engineering;*

- *geometric and performance requirements for components that are not in the scope of separate ISO technical committees.*

Structure

The committee has 9 Subcommittees. One should be mentioned: SC 2 – *Terminology and harmonization of languages.* However, even the other groups may have activities with an indirect CDICT implication.

Projects with relevance to CDICT

ISO / TC 59 has developed a number of vocabulary standards, including **ISO 6707** – *Building and civil engineering – Vocabulary* (in several parts).

There are, however, potential "CDICT implications" in a number of projects, including, but by no means limited to the following more or less randomly selected ones, bearing in mind geographical factors (climate, etc.), building traditions, traditions relating to the use of buildings (including cooking traditions, etc.), legal requirements, etc.:

- **ISO 3055:1985** – *Kitchen equipment – Coordinating sizes.*
- **ISO 6242:1992** (3 parts) – *Building construction – Expression of users' requirements.*
- **ISO/AWI 21930** – *Building construction – Sustainable building – Environmental declaration of building products.*

2.3.5 ISO / TC 68 – *Banking, securities and other financial services*

This committee, as well as some other ISO committees dealing with economic aspects (e.g. TC 222 *Personal financial planning*), have some CDICT relevance, although this is normally embedded in the various projects. We give an even briefer description of this committee that the previous one. This does not, however, necessarily reflect the relative importance and CDICT relevance.

Fields with relevance to CDICT

- Biometric identification techniques.
- Personal identification numbers.
- Privacy of the individual.

2.4 ITU – *International Telecommunication Union*

Home page of ITU: http://www.itu.int/. Home page of ITU-T, the standardization study group: http://www.itu.int/ITU-T/.

The ITU is an international organization within the United Nations system.

All of the activities of ITU fall under ICT, and CDICT issues are touched upon by a number of ITU activities and resolutions. However, it is within the standardization study groups of ITU-T that most of the relevant activities are centred.

Mission

ITU-T's mission is to ensure an efficient and on-time production of high quality standards covering all fields of telecommunications.

Structure

The 14 study groups of ITU-T cover a range of topics relating to the functioning of telecommunication equipment and services. Topics covered include numbering systems, multimedia services and systems, network and service operation, tariff and accounting principles, telecommunications network management systems, signalling, transmission and transport systems, data networks, and value-added services such as universal international freephone numbers.

In order to focus on particular areas of difficulty, study groups create working parties to study specific subjects.

When topics span the mandate of several study groups, a lead study group whose role it is to facilitate the coordinated development of recommendations coordinates the work.

ITU-T Recommendations are issued in 25 series identified by the letters A through Z (no W).

- A – Organization of the work of ITU-T.
- B – Means of expression: definitions, symbols, classification.
- C – General telecommunication statistics.
- D – General tariff principles.
- E – Overall network operation, telephone service, service operation and human factors.
- F – Non-telephone telecommunication services.
- G – Transmission systems and media, digital systems and networks.
- H – Audiovisual and multimedia systems.
- I – Integrated services digital network.
- J – Cable networks and transmission of television, sound programme and other multimedia signals.
- K – Protection against interference.
- L – Construction, installation and protection of cables and other elements of outside plant.
- M – TMN and network maintenance: international transmission systems, telephone circuits, telegraphy, facsimile and leased circuits.
- N – Maintenance: international sound programme and television transmission circuits.
- O – Specifications of measuring equipment.
- P – Telephone transmission quality, telephone installations, local line networks.
- Q – Switching and signalling.
- R –Telegraph transmission.
- S –Telegraph services terminal equipment.
- T – Terminals for telematic services.
- U – Telegraph switching.
- V – Data communication over the telephone network.
- X – Data networks and open system communications.
- Y – Global information infrastructure and Internet protocol aspects.
- Z – Languages and general software aspects for telecommunication system.

(By “languages” in series Z only “artificial languages” are intended.)

There are ITU-T Recommendations specifying the cooperation with ISO, IEC, and ISO/IEC JTC 1.

Projects with relevance to CDICT

Numbering schemes of a variety of types originating from ITU includes much information about nations and regions. All this information seems to be derived directly from national sources. In the rendering of such information in Romanized or “simplified” written forms there are some issues with CDICT relevance.

There are separate terminology recommendations for each of the series of ITU-T Recommendations (these are not found in the B Series, as might be expected).

Some recommendations with CDICT relevance include:

- Tariff principles are covered by several recommendations in the D Series. There are several implications for eBusiness applications.
- Recommendation **E.104 (02/95)** – *International telephone directory assistance service and public access.*
- Recommendation **E.114 (11/88)** – *Supply of lists of subscribers (directories and other means).*
- Recommendation **E.115 (02/95)** – *Computerized directory assistance.*
- Recommendation **E.121 (07/96)** – *Pictograms, symbols and icons to assist users of the telephone service.*
- Recommendation **E.122 (11/88)** – *Measures to reduce customer difficulties in the international telephone service.*
- Recommendation **E.123 (02/01)** – *Notation for national and international telephone numbers, e-mail addresses and Web addresses.*
- Recommendation **E.126 (11/88)** – *Harmonization of the general information pages of the telephone directories published by administrations.*
- Recommendation **E.127 (11/88)** – *Pages in the telephone directory intended for foreign visitors.*

- Recommendation **E.135 (10/95)** – *Human factors aspects of public telecommunication terminals for people with disabilities.*
- Recommendation **E.141 (03/93)** – *Instructions for operators on the operator-assisted international telephone service.*
- Recommendation **E.161 (02/01)** – *Arrangement of digits, letters and symbols on telephones and other devices that can be used for gaining access to a telephone network.*
- Recommendation **E.164 (05/97)** – *The international public telecommunication numbering plan.*
- Recommendation **E.183 (03/98)** – *Guiding principles for telephone announcements.*
- Recommendation **Z.100 (11/99)** – *Specification and description language (SDL).*
- Recommendation **Z.314 (11/88)** – *The character set and basic elements.*
- Recommendation **Z.317 (11/88)** – *Man-machine dialogue procedures.*
- Recommendation **Z.323 (11/88)** – *Man-machine interaction.*

2.5 ETSI – *European Telecommunications Standards Institute*

Home page: http://www.etsi.org/.

Of particular interest is:

2.5.1 ETSI HF (Human Factors): Cultural Diversity and Assistive Technology

This section provides a prime example of how several bodies cover a similar range of standardization topics.

ETSI is primarily concerned with telecommunications, though communications technology and computer technology are increasingly converging. ETSI has a similar general relationship to ITU-T as does CEN to ISO, and some ITU-T Study Groups are also working in similar areas to ETSI. There is some collaboration between ITU-T and ISO in these areas.

EUR 1.4m has been made available from the European Union's eEurope Programme for ETSI HF to set up STFs (Special Task Forces, essentially project teams).

Mobile telephone technology increasingly dominates developments, as does the interaction of mobile telephony with the Internet and the World Wide Web.

The ETSI HF (Human Factors) standardization group is the group in ETSI most concerned with Cultural Diversity in ICT systems.

ETSI HF already has a special mandate for meeting the needs of the Disabled and Elderly, (which fitted in with the earlier TIDE work in the EU). This is also known as Assistive Technology. Outside of ETSI, the ICT Standards Board/ICTSB Coordination Group (covering ETSI, CEN and CENELEC) has been set up to ensure that there is effective collaboration – CEN/ISSS has a new Workshop on Design for All, and CENELEC is working on smart house accessibility.

ETSI HF now has a Special Task Force covering cultural diversity. Initial work revolves around the placement of special letters (additional to abcdefghijklmnopqrstuvwxyz) on mobile telephones. Currently work is underway on additional letters for the Nordic countries and their national languages.

In principle there is no reason why this could not also be extended to cover other languages, and other scripts, just as is the case with alternative layouts on QWERTY/AZERTY/QWERTZ keyboards etc., which ISO/IEC JTC 1 / SC 35 is covering.

Industry input is very strong, involving Nokia, Ericsson, and Siemens in the team with additional input, through inter-industry relationships between Ericsson and Sony, and also Alcatel, Philips, and Motorola.

Partly because of this, within ETSI HF, industry consensus tended to be developed, giving best practice, rather than developing formal standards, which were often seen (by users/developers) as imposed. This allowed specifications to be developed more quickly.

Another ETSI HF STF covered the UCI identifier, which covered different ways of addressing in different countries, and which would cover URLs, telephone addresses, etc. Work on this could also be relevant to work in the W3C Internationalization Workshop on IRIs, which essentially related to extending URLs.

Further funded work covered Icons and symbols (which could be visual, aural or tactile). Some aspects of this are similar to work going on in ISO/IEC JTC 1 / SC 35.

Relevant information on these could be found at http://www.etsi.org/.

2.6 LISA – *Localisation Industry Standard Association*

Home page: http://www.lisa.org/.

SGML mark-up systems (and by extension, XML mark-up systems) are defined in Document Type Definition files (DTDs). Several industries have standardized on various DTDs for the different types of documents that they share.

The standardization work of several bodies, especially some industry consortia, tends to be limited to DTD definition to maximise interoperability for very specific application areas.

LISA is an industry consortium that covers localization standards, and Translation standards, based around XML DTDs for specific use in localization and in translation systems, and also in terminology.

Examples are the TMX and TBX standards (respectively the Translation Memory eXchange Format and the draft TermBase eXchange Format), which are used in machine translation and in localization activities.

OSCAR (Open Standards for Container/Content Allowing Re-use) is the LISA Special Interest Group responsible for their definition of TMX and TBX.

OSCAR also works closely with SCs and WGs of ISO / TC 37 in developing TMX and TBX.

Similar industry/DTD standards are MARTIF and GENETER – again closely developed through close cooperation between ISO / TC 37 and OSCAR.

Some of these are also full ISO standards developed by ISO / TC 37.

2.7 ANEC – *European Association for the Co-ordination of Consumer Representation in Standardisation*

Home page: http://www.anec.org/.

ANEC was established in 1995 to give the opportunity to consumers to be heard in the process of technical standardization. ANEC represents consumers from all European Union and European Free Trade Area countries. ANEC does not run separate standardization projects, but is involved in a number of CEN, CENELEC, ETSI, and ISO projects of interest to the general public.

2.8 Unicode Consortium and the Unicode Technical Committee

Home page: http://www.unicode.org/.

2.8.1 Committees and Consortia: ISO/IEC JTC 1 and the Unicode Consortium

Since 1992, the Unicode Consortium, together with ISO/IEC JTC 1 / SC 2 / WG 2 (*Character sets*) have been jointly responsible for developing the UCS (ISO/IEC 10646-1:2000 and ISO/IEC 10646-2:2001 in its ISO form) and Unicode version 3.0 (with later updates 3.1 and 3.2) in its Unicode form. Essentially these are in many ways the same standard, at least as far as character code values are is concerned.

However, the Unicode Consortium also covers various other aspects related to CDICT in its various related standards, and covers more aspects, and in effect drives the process more than does JTC 1 / SC 2, so both committees and standards are dealt with below, with an emphasis on Unicode.

The Unicode Standard (TUS, itself developed in synchronization with ISO/IEC 10646) is the base standard, accompanied by various technical reports. This is now well standardized, with characters available from all scripts used in official languages of the world, and others being added in a well established process jointly, in effect, with ISO/IEC JTC 1 / SC 2 / WG 2.

Unlike ISO/IEC JTC 1, which separates out work in several different standards, The Unicode Standard also standardizes other aspects besides character codes, like universal transformation formats (UTFs) normalization, ordering, directionality etc.

Fonts are not standardized, that being left up to implementers. An earlier font standardization process, for providing font resources, was AFII (Association for Font Information Interchange) though this proved less useful than had been thought, and has now been closed down, though it was at one stage administered by Unicode.

The latest version (version 3.2) of TUS is also issued as a Unicode technical report (http://www.unicode.org/unicode/reports/tr28/).

TUC also works very closely with W3C in ensuring that the very widely used Web developments such as HTML and XML work fully with Unicode, and also to enable backwards compatibility with existing de jure and de facto standards.

The character code tables of ISO/IEC 10646-1:2000 and ISO/IEC 10646-2:2001 are synchronized with TUS, though the latest version of ISO/IEC 10646 is always behind that of TUS.

Partly because national standards bodies have very formal liaisons, which generally relate only to committees of ISO and ISO/IEC JTC 1, WG 2 tends to provide a wider international input through ISO/IEC JTC 1 / SC 2 / WG 2 structures than is possible through the Unicode Technical Committee alone, though various national standards bodies, governments and companies also participate in TUC.

2.8.2 Unicode documentation

Unicode documentation is of various types: knowing the basic structure is helpful in navigating the mass of standards documentation issued by the Unicode Consortium, as some of these form an integral part of The Unicode Standard. (http://www.unicode.org/unicode/reports/)

In order of centrality to developers and users, these are subdivided into UTSs, UAXs, and UTRs.

A Unicode Technical Standard (UTS) is an independent specification. Each UTS specifies a base version of the Unicode Standard. Unicode 3.0[1] was the latest full publication.

A Unicode Standard Annex (UAX) forms an integral part of the Unicode Standard, but is published as a separate document. New versions of The Unicode Standard are also issued initially as UAXs, prior to consolidation; UAX #27: Unicode 3.1 and more recently UAX #28: Unicode 3.2.

It is expected that Unicode 4.0 will be produced as a cumulative new volume in due course. Other UAXs form an integral part of the main documentation, and include:

- UAX #15: Unicode Normalization Forms
- UAX #9: The Bidirectional Algorithm
- UAX #11: East Asian Width
- UAX #13: Unicode Newline Guidelines
- UAX #14: Line Breaking Properties
- UAX #19: UTF-32
- UAX #21: Case Mappings

In addition,

- UTS #6 (A Standard Compression Scheme for Unicode) and
- UTS #10 (Unicode Collation Algorithm)

are also normative, and though they are separate, they are designated as UTSs.

Unicode Technical Reports (UTRs) was the initial general designation for all Unicode documents. Some normative documents and earlier versions of The Unicode Standard are still UTRs. Nowadays, however, most normative Unicode documents are UAXs and UTSs, and newer UTRs are informative rather than normative.

[1] *The Unicode Standard Version 3.0.* 2000. The Unicode Consortium. Principal authors and editors: Joan Aliprand, Julie Allen, Joe Becker, Mark Davis, Michael Everson, Asmus Freytag, John Jenkins, Mike Ksar, Rick McGowan, Lisa Moore, Michel Suignard, Ken Whistler. Reading, Mass: Addison-Wesley. ISBN 0-201-61633-5.

UTRs include, but are not limited to, reports on: normalizing Unicode text for comparison and storage, compressing Unicode text, mapping to legacy encodings, collating (sorting) strings, linebreaking text, performing uppercase, lowercase, titlecase, and case folding operations, handling CRLF conversions, designing regular expressions, etc.

Some of these are now also UAXs. Examples of UTRs (also including Draft UTRs (DUTR) and Proposed Draft UTRs (PDUTR)) include the following:

- UTR #16: UTF-EBCDIC
- UTR #17: Character Encoding Model
- UTR #18: Unicode Regular Expression Guidelines
- UTR #20: Unicode in XML and other Markup Languages
- UTR #22: Character Mapping Tables
- UTR #24: Script Names
- DUTR #26: Compatibility Encoding Scheme for UTF-16: 8-Bit (CESU-8)
- PDUTR #25: Unicode Support for Mathematics
- PDUTR #29: Text Boundaries

2.8.3 Structures for Unicode and ISO/IEC 10646

The Unicode group of standards as a whole provides normative standards that are essential for implementers to deal with, and pass through the same Unicode Technical Committee. They also provide a great deal of specifications that developers can use than do the equivalent ISO standards, and are regularly supported and updated by industry.

By comparison, in ISO/IEC JTC 1, various standards (like ISO/IEC 10646-1:2000, ISO/IEC 10646-2:2001 and ISO/IEC 14651 are worked on in various committees of ISO/IEC JTC 1. CLAUI itself, which was set up by ISO/IEC JTC 1 to coordinate standards for cultural diversity in ICT, presents a less focused structure than the Unicode Technical Committee.

Nevertheless, committees of CLAUI, in particular ISO/IEC JTC 1 / SC 2 / WG 2, has provided a major opportunity for international input from many countries, which (at least in its earlier days) seemed less evident in the Unicode Technical Committee, which was dominated much more by ICT industry participants from the USA.

The result has been a great deal of coordinated standardization, which have ensured that ISO/IEC 10646-1:2000 and ISO/IEC 10646-2:2001 are harmonized with TUS, and that ISO/IEC 14651 is harmonized with the Unicode Collation Algorithm.

2.9 W3C – *World Wide Web Consortium*

Home page: http://www.w3.org/. A list of current W3C recommendations and other technical documents can be found at http://www.w3.org/TR/.

W3C is an international consortium, with headquarters in the USA and bases at MIT (Boston, MA, USA), at INRIA (Institut national de recherche en informatique et en automatique), Paris, and Keio University (Japan). There is prominent technical input from Europe: European technical participants active in web i18n include: Misha Wolf (Reuters Ltd, UK); Richard Ishida (Xerox GKLS, UK); and Chris Lilley (Graphics Activity Lead in W3C, France) and Yves Savourel (moved from France to RWS Group, Boulder, CO, USA) and Martin Dürst (Switzerland, currently based in Japan).

W3C has been involved in internationalization standardization as it affects the WWW. Its strategy has been

- to use as they stand those base standards which are widely used by industry (Unicode, etc.), typically used in operating systems, or at a base API level;
- in addition to develop particular application level standards that improve ease and consistency of development and use on the World Wide Web (WWW).

2.9.1 W3C Standards related to Unicode

Some base documents are developed together, some as profiles of the other. For example, Unicode can be said to be a profile of ISO/IEC 10646; the W3C Character Model is in many ways a profile of that defined in the UCS

(in ISO/IEC 10646-1:2000 and Unicode version 3.0), and the Unicode Collation Algorithm is a profile of the collation standard ISO/IEC 14651, of which a further profile is the European Ordering Rules (ENV 13710).

The W3C character model allows consistent, interoperable text manipulation on the Web, and a more international Web, by allowing Web documents authored in the world's scripts (and on different platforms) to be written, exchanged, read, and searched by Web users around the world, in all scripts and languages which the UCS repertoire offers, and to provide for cultural conventions.

The W3C character model covers encoding, identification, normalization, string identity matching, string indexing, and URI conventions (URLs and IRIs), the latter providing aliases in all UCS characters for URLs which are restricted to 7-bit ASCII.

Future drafts are likely to cover collation (sorting), fuzzy matching and language tagging.

Various normalization forms are allowed for character representation, but NFC (Normalization Form C) is expected for most purposes in many applications, particularly web-based applications.

Similarly, various Universal Transformation Formats are allowed for character representation, but UTF-8 (Universal Transformation Format-8) is by far the most common in Internet use.

2.9.2 W3C Standards related to mark-up languages

HTML has been the major standard developed by W3C, and W3C retains complete control over this – it is not an ISO standard, unlike SGML (Standard Generalized Markup Language), which is a mark-up language, maintained by ISO.

The use of HTML completely dominates WWW applications, and other mark-up languages have also developed as a result, in particular XML[2]. XML, CSS (Cascading Style Sheets) CGI, PHP, ECMAScript/JavaScript etc., are also widely used in conjunction with HTML, and are also not standardized within ISO.

The Extensible Markup Language (XML) is a subset of SGML. Its goal is to enable generic SGML to be served, received, and processed on the Web in the way that is now possible with HTML. XML has been designed for ease of implementation and for interoperability with both SGML and HTML.

The W3C recommendation (in effect a W3C standard) specifies a syntax created by subsetting an existing, widely used international text processing standard (Standard Generalized Markup Language, ISO 8879:1986 as amended and corrected) for use on the World Wide Web.

The English version of this specification is the only normative version, though translations can be found at http://www.w3.org/XML/#trans.

Details of other W3C XML activities can be found at http://www.w3.org/XML/.

Various standard profiles of XML have been developed for specific purposes, notably DTDs etc. for specific applications. Several of these are also localization-related standards, such as the Translation Memory eXchange format (TMX), and the XML Localisation Interchange File Format (XLIFF) maintained by OASIS, and OpenTag.

One particular feature of WWW applications is that unlike some other applications, where base levels, programming language levels and application levels are clearly separated, mixing program code and user interface is, by contrast, the de facto standard for HTML, as Lloyd Honomichl of Lionbridge pointed out at the IUC21 conference[3].

As a result, it is not uncommon to see a Java Servlet writing HTML code, which in turns contains JavaScript.

However, it may be that in specifying standards, separating the user interface, the program code, and resource files, might be necessary at the standards development level.

Web applications may also have the added complexity of trying to present information to multiple clients in multiple locales simultaneously, as well as dealing with the above complications.

---

[2] *Extensible Markup Language (XML) 1.0* (Second Edition): W3C Recommendation 6 October 2000. Latest version: http://www.w3.org/TR/REC-xml.html. Editors: Tim Bray, Jean Paoli, C. M. Sperberg-McQueen, Eve Maler. – This "second edition" is not a new version of XML (first published 10 February 1998); it merely incorporates the changes dictated by the first-edition errata.

[3] Honomichl, Lloyd: *Nuts & Bolts of Web Internationalization.* (IUC21, Dublin, May 2002).

2.10 Other industrial consortia and similar initiatives

2.10.1 IETF and IANA and SIL International

IETF – the Internet Engineering Task Force – http://www.ietf.org/.

IANA – the Internet Assigned Names Authority – http://www.iana.org/.

SIL International (formerly known as the Summer Institute of Linguistics) – http://www.sil.org/ – has as its purpose to work with language communities worldwide to facilitate language-based development through research, translation, and literacy.

2.10.1.1 Language codes and country codes

Particularly in XML implementations, base standards such as ISO 3166 and ISO 639 are also important in a variety of web transactions.

Various RFCs (Requests for Comment – many regarded as actual standards) have been issued under the auspices of the Internet Engineering Task Force (IETF). These also affect the web, and also use existing ISO standards and W3C standards as base standards.

In terms of Cultural Diversity in ICT, RFC 3066 (Language Tags) is particularly important. This extends the earlier RFC 1766 (Language Tags) which used approximately 300 alpha-2 identifiers from ISO 639, by also allowing the inclusion of (approximately) an additional 100 alpha-3 identifiers from ISO 639-2.

IANA also has a Language Tags Reviewer, who can add further language tags and sub-tags (as allowed for in IETF's RFC 3066), where the current range is insufficient.

However, the World Wide Web includes web resources in far more languages than any of the above can currently cope with, so there is something of a bottleneck at this point, with supply not meeting demand.

There is a widely used scheme of 3-letter tags developed over many years by SIL, which is in widespread use among linguists, but as a separate scheme.

Some incorporation of this, or something like it from another quarter, looks like a way of meeting this problem, and SIL is now involved in discussions with ISO / TC 37 / SC 2.

2.10.1.2 Character sets – IANA

In relation to the WWW, IANA has taken on similar role to ECMA and ISO, in maintaining a register of character sets.

When using web browsers, sometimes there is distortion or loss of data, when information on the web is presented in a different character set. Sometimes solutions are at hand in the browser to correct this, but sometimes they are not.

UCS (ISO/IEC 10646 and Unicode) is increasingly the expected character set on the World Wide Web.

However, much legacy data is available using a wide variety of other character sets.

Just as in earlier days, ISO 2022 and ISO 2375, and the associated International directory permitted some languages and scripts to be used at all in ICT systems, IANA uses a similar approach to enable the development of mappings between various non-UCS character sets and UCS, so that information can be presented correctly, without distortion or loss of data.

Software that will make all possible necessary changes on the fly might be expected to be more widely available in due course, but such software needs its data from some source.

IANA provides this by maintaining a register of character sets, larger than the original International register, and also including vendor-specific character sets.

The IANA registry (http://www.iana.org/assignments/character-sets) gives a unique "MIBenum" numerical identifier, and also one or more "aliases" that can be used in identifying these character sets in any information interchange.

Example:

Name: ANSI_X3.4-1968
MIBenum: 3
Source: ECMA registry
Alias: iso-ir-6
Alias: ANSI_X3.4-1986
Alias: ISO_646.irv:1991
Alias: ASCII
Alias: ISO646-US
Alias: US-ASCII (preferred MIME name)
Alias: us
Alias: IBM367
Alias: cp367
Alias: csASCII

2.10.2 ECMA – *Standardizing Information and Communication Systems*

Home page: http://www.ecma.ch/.

ECMA is an international industry association founded in 1961 and dedicated to the standardization of information and communication systems.

Specialist Technical Committees are established to perform particular tasks, and are disbanded when the task is completed. For efficiency, much of the work of Technical Committees takes place electronically, with formal meetings only taking place a few times a year. The current active Technical Committees within ECMA are listed below: links give access to the full Scope, Programme of Work and Membership. Specific files illustrating the work in progress are available for some groups.

- TC12 – *Product safety*
- TC15 – *Volume and file structure*
- TC17 – *Magnetic tapes and cartridges*
- TC20 – Electromagnetic compatibility
- TC26 – *Acoustics*
- TC31 – *Optical disk cartridges*
- TC32 – *Communication, networks and systems interconnection*
- TC32-TG11 – *Computer Supported Telecommunication Applications (CSTA)*
- TC32-TG14 – *Private integrated services / Corporate networks – Services and signalling*
- TC32-TG17 – *IP-based multimedia communications*
- TC34 – *Office equipment*
- TC36 – *IT security*
- TC38 – *Product-related environmental attributes*
- TC39 – *Programming and Scripting languages*

ECMA publishes Standards and Technical Reports. Some 200 Standards and 30 Technical Reports are currently valid.

Projects with relevance to CDICT

One ECMA Standard deals with 7-bit character set, and ten Standards are dedicated to 8-bit character sets. The most recent are updated in December 2000.

- **ECMA-262** – *ECMAScript Language Specification,* 3rd edition (December 1999)
- **ECMA-290** – *ECMAScript Components Specification* (June 1999)
- **ECMA-327** – *ECMAScript 3rd Edition Compact Profile* (June 2001)
- **ECMA TR/53** – *Handling of Bi-Directional Texts,* 2nd edition (June 1992)

2.11 Other governmental and intergovernmental initiatives

Some intergovernmental organizations (IGOs) are also involved in standardization, particularly in areas that have an impact on legal and financial issues. In the most successful cases these also work closely with committees of either ISO/IEC JTC 1 or ISO.

As an example, the International Civil Aviation Organization's Technical Advisory Group (ICAO TAG) is the driving force behind ISO/IEC 7501 (*Machine readable travel documents*) which is produced by ISO/IEC JTC 1 / SC 17 / WG 3 (*Identification cards – Machine readable travel documents*). This takes an official ICAO document as its base document.

Similarly, UN/CEFACT in Geneva works closely with ISO/IEC JTC 1 / SC 32 (Data Management Services) in developing various standards related to EDI (Electronic Data Interchange).

UN/CEFACT has also produced, and maintains, the LOCODES (Location Codes) standard, increasingly used in EDI transactions, which is not an ISO standard, but which is widely used, and expands the current 3-letter Airport Codes standard to cover many more towns and ports in many countries.

A third example is the United Nations Statistical Office in New York, whose data on country names and numerical codes are used normatively in the country code standard ISO 3166 (although ISO 3166 presents only a subset of UN data in its tables, not using codes for regions for example, unlike the UN data).

A fourth example is the International Telecommunications Union's ITU-T Study Groups, some of which have been concerned with character sets. ITU-T also liaises strongly with ETSI (see 2.5).

A final example is due to both the International Labour Organization (ILO) and Unesco being active in developing early cooperative information retrieval systems. This began in the 1960s, but these systems, particularly ISIS and CDS/ISIS, are both still very widely used (mainly in scientific institutions) on a global basis, and developed on a continuing basis.

As a result, both ILO and Unesco were also involved in developing and maintaining mainly bibliographic data element standards for these, which are incorporated in other standards, notably ISO 2709, the UNISIST Reference manual, and in UNIMARC.

3 Scope of Cultural Diversity in ICT: inclusions, exclusions

There is a need to identify the boundaries of standardization for Cultural Diversity in ICT, and also to identify the basic themes within it. Much of this is already done, e.g. in the PWC report, but it will also be necessary to identify future trends etc., and interrelationships between the themes.

It will be necessary to determine the scope, especially inclusions, exclusions of work.

Three existing taxonomies that relate to the work in this area have been examined in looking at CDICT scope:

(a) from the PWC report;

(b) from earlier taxonomy work in CEN / TC 304;

(c) from the eEurope Action Plan.

All have their strengths and limitations in this area, though it should be pointed out that (c) was not designed with CDICT at the forefront.

(a) The PWC report noted the following areas:

- Infrastructure aspects (fixed and mobile telecommunications infrastructures and connection policies, network protocol and mark-up language internationalization);
- Input/output aspects (keyboards, character sets, ordering, sorting, data formats);
- Linguistic aspects (translation, writing style);
- Design and content aspects (look-and-feel, localized content);
- Commercial aspects (branding, marketing, communication, client services, pricing, payment mechanisms);
- Legal aspects (applicable law, consumer protection, privacy, liability, complaints)

The earlier drafts of the current PT report also proposed analysing these down into the following sections:

- Telecommunications infrastructures and connection policies
- Network protocols
- Conformance and testing
- Data formats and metadata
- Electronic commerce
- Mark-up language internationalization
- Character sets
- Alphabetical ordering
- Character matching
- Keyboards
- Language identification
- Multilingualism
- Translation
- Writing style
- Speech applications
- Language resource management
- Terminology work (including Administrative nomenclature)
- Individual languages, language groups, and regions

We propose to include these items when recommending the allocation of work. However, we propose to regroup them under taxonomy (c) – the eEurope Action Plan, and to conflate, rename or abandon some (for example, Writing style may be better regarded as Orthography, and listed under some aspects of locales).

(b) Some earlier work was done by Keld Simonsen for on taxonomy. This provided a lot of detail – perhaps more than a standards committee needs – and its suggested allocation of work among ISO and CEN committees also proved controversial. Nevertheless, this document will be analysed, and if there is anything useful that is not already in (a) or (c), this will also be incorporated.

(c) The eEurope Action Plan also has a taxonomy of areas of work, several of which relate to Cultural Diversity in ICT systems. The table below is conflated from the 2002 and 2005 eEurope Action Plans (using the headings from the 2002 Action Plan).

1. A cheaper, faster, secure Internet
   a) [infrastructure] Cheaper and faster Internet access
   b) eResearch Faster Internet for researchers and students
   c) eSecurity Secure networks and smart cards
2. Investing in people and skills
   a) eLearning European youth into the digital age (eEducation)
   b) eWorking Working in the knowledge-based economy
   c) eAccessibility Participation for all in the knowledge-based economy
3. Stimulate the use of the Internet
   a) eCommerce Accelerating eCommerce
   b) eGovernment Government online: electronic access to public services
   c) eHealth Health online
   d) eContent European digital content for global networks
   e) eTransport Intelligent transport systems

CDICT affects items in Group 3 most strongly. Work on CDICT in items in Group 3 can also be used in developing CDICT parts of the specifications in Group 2 (and in 1(b)). Group 1, apart from that, is less heavily affected, as it relates to system to system interfaces, rather than system to user interfaces, although there will be aspects that need to be taken care of.

These three areas will be explored in greater depth in chapter 4, where we look at allocating work in CDICT within CEN/ISSS.

It should be assumed that any standards assume certain specific international base standards to be already included.

Standards relating to UCS (ISO/IEC 10646 and Unicode) should be expected as a norm, and less work should be done which involves earlier seven and eight bit codes.

There should also be support for Europe-specific standards for XML implementations.

It might be expected that other European standards might be needed to cope with needs in specific areas, e.g. for the Nordic countries, or for the Baltic countries – specific needs for Sami languages being an example.

There should be a focus on standardization for future needs, rather than on legacy needs. The likelihood is that somebody will be developing solutions already for legacy needs, where they need to be developed.

Standardization should cover specifically European needs that are not being met at the international level. This may mean developing European standards which subset wider (and larger) international specifications.

No European standards should duplicate work done at the international level.

Most of the central operations that are affected by Cultural Diversity use (a) pre-defined processes that use (b) data from one or more tables.

Typically, such standards define both the processes and the contents of those tables. Examples of this include character sets, character repertoires; ordering/sorting; use of identifiers for date, time, currencies, languages, countries; subdivisions of any of the above; locales (combinations of the above); terminology; Metadata specifications; XML specifications; data conversion (including transliteration, fallback); translation elements; etc.

All of the above examples use both table(s) of data and define relevant processes for their use in IT systems.

In some cases additional standards build on or “inherit” values from standards that define a more basic level. Examples might include Unicode standards adding further features to the basic coded character set elements in ISO/IEC 10646; RFC 3066 adding further language identifiers to the codes in ISO 639-1 and ISO 639-2; locale specifications combining elements from various standards; HTML building on SGML; XML building on SGML; various industry-specific XML implementations building on XML specifications, etc.
In some cases some proprietary practices become so widely spread through commercial success that they become a de facto standard, and may be incorporated into existing standards. One example is Digital’s “Roman-8” character set becoming the basis of ISO 8859; IBM’s use of EBCDIC in mainframes; another is the widespread use of Google, and its matching capability, which seems to have arrived at some of the same matching principles recommended by the Browsing and Matching Project Team set up by CEN / TC 304.

Similarly, some standards are so basic that they are built into operating systems, and/or APIs, so that application developers do not need to provide additional capability in their own developments, but can merely “inherit”

them, as in character set or ordering. However, sometimes the application will need to add additional things, where either the underlying specification provides insufficient detail, or a different process needs to be added to produce different results.

In the core standards, in relation to a specific area, there is also (i) a need to define both whole data-sets for possible global use, while (ii) many users only require restricted sets for use in a particular area or situation, as in the cases listed below.

4 Developing a programme, and allocating work

As noted in chapter 3, the eEurope Action Plan also has a taxonomy, or schema, of areas of work. In several of these elements, the issue of Cultural Diversity in ICT systems needs to be taken care of, in order to make them work effectively.

Note that there is a particular emphasis on eGovernment in this section, as most countries in Europe have an increasing emphasis in this area, and also because CDICT issues are relevant in so many areas of this.

It is also important that any standards that involve CDICT can also be taken up in real world situations, which will be assisted if various standards activities related to eEurope can be enabled to interface with each other.

A schema for such interrelated activities is shown in the table below. The table is based on that in chapter 3 of this Project Team report, which itself is based on a schema in the eEurope 2002 Action Plan. In addition we also add new section headings within chapter 4 to this table.

It may be that CEN/ISSS might follow some of the above by setting up new Workshops or other bodies. Alternatively it might wish to liaise with bodies that already exist in this schema within a European context. In any case, if any of this were to be adopted, it would need to be done in an evolutionary way over some time: such changes could not take place overnight.

In any case, this section notes that most of these areas have CDICT elements that need to be addressed in standards implementation and use in those areas.

Contents of chapter 4

Notes

Clause 4.1 provides an overview, while clauses 4.2–4.12 describe individual “eElements” associated with the eEurope work, and clause 4.13 proposes additions.

Numbers in brackets are taken from the eEurope Action Plan 2002.

4.1 eElements: what are their CDICT standardization needs?

Individual “eElements” from the eEurope Action Plan are described below, in clause 4.2 onwards.

It is proposed that existing CEN/ISSS Workshops and other activities, such as CEN TCs, be allocated to these headings, and where necessary, other Workshops, and/or TCs, and/or PTs be set up to match these headings.

Where necessary, it will also be necessary to propose liaisons with other relevant bodies, such as ISO or ISO/IEC JTC 1 committees, and organizations of the UN system, and consortia, in ensuring that European needs identified from the above framework, are met.

CEN/ISSS could provide an extremely useful service in helping to provide standards in these areas. Inevitably, domain-specific standards are beginning to be developed by those most involved in a specific eElement, and where this is so, this could be the basis of any further work within CEN/ISSS, much as other Dublin Core activities were used as the basis for work within the European Workshop on Dublin Core and Metadata.

It could be argued that Workshops and Project Teams are the natural way forward for all of these: they naturally lend themselves towards the workshop approach and the project team approach, which would involve experts in the fields concerned, rather than nominated delegates representing National Member Body views.

However, comments from National Member Bodies could be useful, if any national standards organizations are getting involved in these areas, but it should be the involved experts and users who should drive it, rather than it being driven by national standards bodies.

4.2 Infrastructure aspects of eEurope

Much of this relates to providing cheaper and faster Internet access for Europe's citizens, and providing the telecommunications and systems aspects, which will allow other eElements to "inherit" the ability to allow CDICT to operate.

This element does not have much separate CDICT needs, other than to ensure that standards such as ISO/IEC 10646, ISO/IEC 14651, and the related Unicode Standards and specifications, together with XML-related specifications, are naturally available, and that it also allows for backwards compatibility with other systems elements which interact with the European and global networks, so that any information is not distorted by limitations of the users' equipment.

These type of concerns are in any case typically already in place as infrastructure is developed and added to, and are already taking shape in Europe.

By the end of 2003, the Commission plans in any case to issue an agreed interoperability framework to support the delivery of pan-European eGovernment services to citizens and enterprises. See the following references:

- OJ C 292 of 18.10.2001.
- OJ C 86 of 10.04.2002.
- http://europa.eu.int/information_society/eeurope/egovconf/index_en.htm
- Information society technologies (http://www.cordis.lu/ist/).

In addition, the European Commission is preparing an initiative to follow up the Communication on creating a EU framework for the exploitation of public sector information (see COM (2001) 607 final of 23.10.2001).

4.3 eResearch

In the EU 2002 Action Plan, the work on eResearch is described as providing "Faster Internet for researchers and students." Much of that relates to infrastructure issues, and this is discussed further in clause 4.2.

Content and delivery are also major issues, which apply equally to eEducation (aimed at school-age learners), and eWorking (aimed at education and training within the workplace). For that reason, eResearch, eEducation and eWorking are all dealt with together in clause 4.5 eLearning.

4.4 eSecurity

eSecurity deals mainly with infrastructural issues, in most transactions taking place between systems, rather than between systems and users.

However, it is also possible that some issues of CDICT may be relevant to issues of security and electronic authentication systems, particularly in matters of personal names, although again CDICT will not be as central to eSecurity as it is in relation to other aspects documented in clauses 4.3–4.12.

Nevertheless, action on personal names is important: the occasional inability of both people and systems to deal with variations on names has caused work for national administrations, and anxiety for individuals, far beyond what might be expected, as is noted in chapter 2, and these issues can only be solved if CDICT issues are taken into account.

Liaison with committees like ISO/IEC JTC 1 / SC 7 / WG 3, and ICAO, are important in this area.

4.5 eLearning

This deals both with infrastructure, and with content, and can also be said to have three sub-elements: eResearch, eEducation, and eWorking, given that learning and training are increasingly lifelong activities.

Infrastructural aspects of these would be addressed in clause 3.2.

eResearch in the EU builds on existing Internet use for both research and teaching within the higher education community. Developments here have led developments in the eLearning and eWorking areas, and arguably developments here have generally led developments in eEducation and eWorking.

eEducation follows the EU's "eLearning" initiative to adapt the EU's education and training systems to the knowledge economy and digital culture. This is aimed at the fulltime education sector. There are four components, some infrastructural:

- to equip schools with multimedia computers,
- to speed up the networking of schools and teachers.

some related to eWorking:

- to train European teachers in digital technologies,

and some related to eContent:

- to develop European educational services and software.

The last of these has major CDICT issues.

Similar comments can be made about eResearch, which could again be analysed in terms of the same elements, except that infrastructural developments are deeper in place.

4.6 eWorking

eWorking is also analogous to eResearch and eEducation, in also having similar infrastructure and content areas. CDICT elements would also be relevant here as well, and in this case liaison with departments of employment, chambers of commerce, and other work-related bodies would be appropriate.

4.7 eAccessibility

eAccessibility involves making Web content – and indeed ICT systems in general – accessible to people with disabilities. However, some developments also benefit non-disabled users too, e.g. enabling access on a variety of equipment (e.g., desktop browser, voice browser, mobile phone, automobile-based personal computer, or using a hands-free environment, etc.).

In addition, eAccessibility has various analogies and crossovers with eDiversity, and again it makes sense to have a special body – a workshop or otherwise – to cover eAccessibility.

ETSI HF already plays a comparable role in this area, and if a CEN/ISSS Workshop were to be set up in this area, (a) it would be need to be analogous to ETSI HF, and (b) it would need to be in close liason with ETSI HF.

Alternatively, it may be sufficient that CEN/ISSS and ETSI HF can liaise directly with each other, in relation to a schema such as that above, as long as any necessary procedures for harmonisation can be built in.

It has been noted above that Sound applications are likely to play an increasing role in ICT systems, and that there may be a case for having a special body – a workshop or otherwise – to cover what might be called eSound.

However, if CEN/ISSS set up a body to cover eAccessibility, and if many of the people involved are also involved in ETSI HF, and thus also involved in sound applications through their involvement in telecommunications companies, there may be an equal case to state that an eAccessibility group might also cover eSound as part of its scope (see also references to (1f) and (1g) in clause 4.13).

4.8 eCommerce

In order to realise the Internal Market for eCommerce, the eEurope Action Plan proposed to adopt eCommerce legislation and promote self-regulation, establish electronic marketplaces for public procurement and encourage SMEs to "Go Digital".

There is a significant need for CDICT in this action area.

There is already contact between CDSG and the CEN/ISSS eCommerce Workshop, but this needs to be built on substantially, on a continuing basis, particularly if an eDiversity group is set up.

4.9 eGovernment

Standards for eGovernment would cover interactive delivery on-line of government services, including public services such as eHealth or eTransport. In addition, they would cover all levels of government (central, regional, local) and would also address back-office changes that have proved so far most relevant to implement interactive public eServices.

The scope for that is based on the EU's 2002 eGovernment conference http://europa.eu.int/information_society/eeurope/egovconf/about/index_en.htm.

Typical examples include applications for driving licences, registrations of births, marriage and deaths.

It enables better access, reduced costs, and can avoid issues of favouritism and fraud, where there is a risk of that, in some parts of the world. It can particularly reduce costs by allowing greater throughput of transactions between the government and its citizens, and is particularly cost effective at cutting various costs in rural areas.

In addition, eGovernment can also be divided into specific areas, including eHealth and eEducation, which include some of the same aspects as eGovernment, but also have additional needs, e.g. to allow a much higher level of personal interactions between any user and the professional who is providing a service.

For all these cases, there is an urgent need for standards for eGovernment: many different initiatives have been tried across Europe, and there will be a need to pull out best practice and to develop standard procedures. The European Union's IDA programme, which covers Interchange of Data between Administrations (http://europa.eu.int/ispo/ida/), also looks at standards.

Some of the same processes and procedures will apply in eGovernment that also apply in eCommerce, although there will be many differences too. It may be expected that there is considerable crossover here, which any standards body considering eGovernment should also consider.

There is also an urgent need to cater for Cultural Diversity in eGovernment, more so than in eCommerce. As well as needing documentation and assistance in many national languages across Europe, several countries provide documents as a matter of course in minority languages across the countries concerned. There is also a need to provide for local variants of languages: in providing information on legislation and food standards in the potato industry, for example, it is important to emphasise national/local uses, even at as basic a level as describing the potato as Kartoffel in Germany, and Erdapfel in Austria.

There are a lot of examples of multilingual eGovernment outside of Europe, which can be drawn on. eGovernment proves particularly effective in areas that have widely spread rural populations, in cutting down or avoiding the costs of bureaucratic processes, and travel or transport costs associated with these processes.

One example is parts of Canada, where English, French, Inuktitut and Cree, and other minority languages – each with several different dialects – may also be in use. Various eGovernment applications are in use in Ontario and in Quebec.

Another example is in India, where there have also been a lot of eGovernment initiatives, from the various states of India, many of which have multilingual components.

One feature that has emerged very strongly from India is the need for local "gatekeepers", who have a respected foothold in the area concerned, who are able to enable the various transactions for a generally less IT-literate population in many villages. There may also need to be standards for gatekeepers too, within a set of European eGovernment standards.

Certainly some key groups, particularly within rural areas, may be key in enabling the take-up of eGovernment. Farmers (if they can see an economic benefit) may be one such key group in rural areas. Including some eCommerce aspects, such as information on market prices for agricultural commodities, in local eGovernment centres, also increases take-up of eGovernment services.

It will also be necessary to work with farmers' organizations, and various other interest groups, who are in effect user organizations, in rural areas in order to enable standard approaches. A similar set of approaches needs to be adapted in urban areas too, in order to increase take-up.

Some aspects of standardization for eGovernment and for eCommerce need to be taken together, as there are many overlaps in approaches, and in the technologies involved, as well as many unique features to each.

Initiatives regarding eGovernment exist at national, regional and local level. These are complemented by actions carried out at European level, several of which have CDICT aspects.

The EU's IDA programme supports interoperability, standardization and the provision of pan-European services, and the IDA portal now being developed plans to offer multilingual access points for online information and services to citizens and enterprises supporting mobility across borders.

It is also planned that the IDA portal should serve as test bed for multi-platform technologies, including the use and improvement of machine translation systems to fully cover the linguistic diversity of European users of such web sites.

There is a danger in such plans that only the largest languages are catered for, as is the case for existing commercial MT systems used on many websites. It would be useful for CEN/ISSS to be able to provide guidelines on developing such services to cover other languages in use within Europe as well, both with national and local official status.

4.10 eHealth

This also deals both with infrastructure, including data networks, and with content, including health information and legal information.

The EU's Health Online plan envisages the implementation of an infrastructure that will provide user-friendly, validated and interoperable systems for medical care, disease prevention, and health education through national and regional networks that connect citizens, practitioners and authorities on-line.

There are a lot of doctor-to-patient interactions and nurse-to-patient interactions, as well as intra-professional interactions, and accuracy of information (itself affected by cultural diversity) is vital, and affects patient outcomes.

Users may describe symptoms in a variety of languages, and various different ways, and building in systems that can cope with cultural diversity is vital, in terms of outcomes.

The practices and experiences of different eHealth initiatives in various member states will also need to be looked at in developing standards for eHealth.

4.11 eContent

eContent is a market-oriented programme that aims to support the production, use and distribution of European digital content and to promote linguistic and cultural diversity on the global networks.

It is the only one of these EU actions which aims to promote linguistic and cultural diversity on the global networks, though this may be limited mainly to Language and customisation players, rather than across all applications. See http://www.cordis.lu/econtent/.

It is assumed that the CEN/ISSS Workshop on Dublin Core and Metadata might fit in here. Alternatively, that could be more relevant in part of the Infrastructure schema.

There are certainly some aspects, such as rights, which are a big part of what commercial players talk about in relation to content, and documenting rights is an important part of the specification of the CEN/ISSS Workshop on Dublin Core and Metadata.

Where that Workshop might go in the schema might depend on what developments in XML were incorporated in infrastructure aspects.

4.12 eTransport

eTransport mainly relates to Traffic & Travel Information Services (TTI) within and between destinations in Europe.

There are significant elements of cultural diversity that arise in this area. These do not just relate to use of language, but to many other cultural elements, including currencies, time zones, countries, areas, and locations.

There are also the results of R&D projects funded by the European Union under its previous Framework Plans, that could also be drawn on (and drawing on the results of such projects could also be fruitful under most of the other eElements above.

This area would be a good test bed for setting up an additional standardization group (workshop or otherwise), as there would be many experts that could be drawn upon from all over Europe, who are working on such systems now, and governments and private companies are all involved, as are European bodies such as the European Conference of Ministers of Transport, and (in the domain of air transport) international bodies such as IMO and ICAO, and NGOs such as IATA – and many others, in many other transport domains.

Indeed, many other application areas may be able to use the outcomes of such a targeted group as examples for their own use.

Alternatively, it would be useful for a group such as this to evaluate some of the specifications produced by other groups as to how well they meet wider uses (or not), with an area such as this providing a good test bed.

4.13 What should be added to the above schema?

Chapter 3 already proposes adding a body (a workshop or otherwise) to cover CDICT, to liaise with all the elements above. It appears to the PT that this might be best placed together with other infrastructural aspects, as a new (1d) in the above picture, and might be best entitled eDiversity.

There are further areas that may each necessitate the setting up of a body within CEN/ISSS, i.e. (a) Machine Translation and (b) Sound applications.

Machine Translation (MT – properly described as Machine Assisted Translation in practice) is increasingly relied upon as a component on the World Wide Web in multilingual situations. Translation is an area which CEN / TC 304 at least did not cover: so it is proposed that a separate body be set up within CEN/ISSS to cover use of translation, rather than just adding it as work within a CDICT body. In terms of the schema above, its ideal placement would be with other infrastructural aspects, e.g. as a new (1e) in the picture above.

For Sound applications, it is likely that these will play a greater role as technology matures, and as it becomes possible to include more text-to-speech and speech-to-text applications on web-based systems. Again, this should be placed with other infrastructural aspects, e.g. as a new (1f) in the picture above.

Finally, it is suggested that eAccessibility or Assistive Technology is as much an infrastructural aspect as are CDICT, Machine Translation and Sound applications, and each of them also have a bearing on the other, so in terms of a schema, this should again be placed with other infrastructural aspects, e.g. as a new (1g) in the picture above.

5 Recommendations

The recommendations are given in summary form, rather than a narrative form. The recommendations were revised following the advice of the Cultural Diversity Steering Group (CDSG), following the CDSG meeting on 12 July 2002. The recommendations cover:

(a) Identifying the programme of activities to support CDICT within Europe. *Note:* Examples are given below, that will be expanded in subsequent sections.

(b) Proposal for structure to cover that work. *Note:* In the text below, a "CDICT Advisory Group for Europe" ("CAGE") is used as a working title and a working acronym. The title and acronym are used for the purpose of reference in this report only.

(c) Where solutions (or other bodies) are already identified, recommending the use of appropriate existing solutions (or other bodies) to enable this within Europe.

(d) Identification of who should perform that work.

(e) Identifying resources, and the timescale for completion of that work.

Recommendation 1

The strategy for activity in this area should be to provide CDICT support for the complete body of standardization work in CEN. It should either

(a) work with existing TCs and/or WSs of CEN, to identify work where CDICT is relevant, and to assist in its development, or

(b) develop underlying CDICT-related specifications that can underlie the work of several TCs or WSs.

For examples and action, see recommendations 2 and 3.

Recommendation 2

The work programme should provide for CDICT aspects for themes in the eEurope 2005 Action Plan, especially underlying specifications.

Example: further work on Matching for European needs.

Action: CAGE.

Recommendation 3

The work programme should also provide for CDICT aspects for themes in the eEurope 2002 Action Plan, especially with TCs and WSs already involved in those themes, in those cases where CDICT issues remain underdeveloped.

Example: maximising use of European LOCODES (Location Codes) in intelligent transport systems under development, for maximum portability and interoperability of systems.

Action: TCs and WSs of CEN, with input from CAGE.

Recommendation 4

The work programme should not develop further base standards, or European profiles of base standards, except as specified in Recommendation 5, as these are already sufficiently documented by organizations such as ISO/IEC JTC1, the Unicode Consortium, and ECMA.

Examples of non-development: ISO/IEC 10646, The Unicode Standard and its Technical Reports series, ISO/IEC 14651, ISO/IEC 9995, POSIX/Linux specifications, ECMAScript/JavaScript , C#, etc. do not need specific European specifications.

Recommendation 5

The only European profiling that should be done, is where documentation for European CDICT aspects is just not available elsewhere.

Example: Completion of descriptions of European matching requirements.

Action: CAGE, or revised Matching PT.

Example: Assistance to ISO / TC 37 / SC 2 and other concerned bodies in identifying European languages which require language identification codes (as was identified earlier by CEN / TC 304).

Action: CAGE, ISO / TC 37 / SC 2, IETF, IANA, SIL.

Example: Providing details of national and other keyboard layouts in Europe to ISO/IEC JTC 1 / SC 35, for planned international register of keyboard layouts.

Action: CAGE, National Member Bodies.

Recommendation 6

The work programme should also include completing the few remaining work items still assigned to CEN / TC 304, before assigning dormant status to CEN / TC 304.

Examples:

- **prCR 14381** – European Fallback Rules – Fallbacks for Latin, Greek and Cyrillic scripts.
- **prCEN/TS 1923rev** – European Character Repertoires and their Coding – 8 bit single byte coding.
- European additions to OCR-B.

Action: CEN / TC 304.

In addition, ensuring that narrative locale specifications for European countries, taking into account their typical languages, are available from, or have received comments from, National Member Bodies.

Action: CEN / TC 304, also soliciting input from CEN member bodies.

Recommendation 7

The work programme should also include a Technology Watch element, given the rapidly changing ICT environment, particularly monitoring Web and XML developments, and other new technologies as they arise. It should regularly assess which may be relevant, and which may be transitory, or limited in application.

Example: monitoring XML systems development to assess implications for CDICT in Europe, comparing eBXML, UBL (Unified Business Language) and CXBL, assessing any gaps relevant to Europe, assessing the likely use of these and other new technologies, and making recommendations to CAGE.

Recommendation 8

The work programme should also include a dissemination element, given the importance of ICT, and CDICT, in all aspects of European life.

Example: organization of conferences/meetings for developers.

Action: CEN.

Example: contact with chambers of commerce, universities, EU projects involving ICT.

Action: CDSG/CAGE.

Example: Maintenance of a European CDICT portal.

Action: CAGE, DIFFUSE Project partners, others.

Recommendation 9

To carry out this work programme, and to enable coordination, we recommend that a CDICT Advisory Group for Europe (CAGE) (working title and working acronym for reference in this report only) be set up, to take over the work of the current Cultural Diversity Steering Group.

Action: CEN to approve CDICT Advisory Group for Europe (CAGE) being set up in place of the Cultural Diversity Steering Group.

Action: CDSG/CAGE and CEN to develop Business Plan for CAGE (see draft Business Plan in chapter 7).

Recommendation 10

Funding for this work programme should be examined. CDICT is a necessary overhead given the multinational and multilingual nature of Europe, and for that reason it needs more than a standard workshop approach, where participants pay for an obvious industry benefit. CDICT is necessary, but its benefits are underlying and diffuse, rather than with an obvious payoff for immediate participants.

It is therefore recommended that CEN examine a mixture of public and private funding for the activities of CAGE. Where possible, public funding should be allied to funding elements in the eEurope 2005 Action Plan.

Example: seeking funding from the European Commission for
(a) any underlying CDICT projects which require a project team, and
(b) any conferences planned on CDICT.

Action: CEN.

Example: seeking commercial sponsorship for technology watch and dissemination parts of the work programme, allowing prominence of the sponsor's name in publications.

Action: CEN, CGSD/CAGE.

Recommendation 11

In terms of timescale, CAGE needs to take account of the rapid takeup of CDICT-related standards, both from ISO/IEC JTC 1 and from consortia, and the much shorter life cycles for standards, and the need to take account of rapidly developed industry standards.

Action: CAGE to review existing projects on a regular basis (perhaps more frequently than the usual 5-year cycle) to assess whether to confirm, update, or deprecate them.

6 Initial work programme for CAGE

This chapter provides an initial work programme for CAGE, and recommends priorities for this point in time. As CAGE develops, this will need to be added to or modified, once work gets under way.

It marks a shift away from developing specific base standards for Europe (and from technical reports and profiles on those base standards), which represents much of the work of CEN / TC 304, towards meeting specific objectives from government and industry in Europe, which were identified in key European Union documents, specifically the eEurope Action Plan for 2005, and additional ones from its predecessor, the eEurope Action Plan for 2002, as well as some identified in the PWC report.

Action: CAGE should develop a programme of work related to ongoing work in
(a) CEN TCs and WSs;
(b) industry consortia;
(c) ISO/IEC JTC 1 and ISO.

This should relate to the work areas identified in
(i) the eEurope Action Plan for 2005,
(ii) the eEurope Action Plan for 2002, and
(iii) additional work areas not covered in the above, identified in the PWC report and/or by the current PT.

The work programme will need to be carefully managed, as the degree of potential overlap is very high between different domains in these areas.

(i) The eEurope Action Plan for 2005 includes the following areas:

- eGovernment
- eLearning
- eHealth

(ii) The eEurope Action Plan for 2002 includes the following areas in addition, where CDICT issues still need to be further developed:

- eAccessibility
- eCommerce
- eContent
- eTransport

(iii) There are also other areas, which may need future work related to CDICT in Europe, identified in the PWC report, and/or by the PT:

- terminology work
- translation
- language identification
- transliteration in passports
- speech applications

Note: Many of these work items will need to be arranged by, with, or with the approval of specific TCs or WSs of CEN. The specific TCs or WSs involved are not indicated at this point, as discussion of this proposed Work Programme by CDSG, CEN/ISSS Forum, and specific TCs and WSs would be necessary. Nevertheless, working with, and where appropriate, under the direction of, specific TCs and WSs is encouraged.

6.1 Work related to themes in the eEurope Action Plan for 2005

6.1.1 eGovernment – Government online: electronic access to public services

Work Item: Government online: electronic access to public services.

Deliverable/Actions:

Review of web-based access across different national governments and EU pages, covering information provision and form-based input. Analysis of web-based and mobile-phone-based developments in CDICT and eAccessibility.

Generalised Best Practice recommendations based on ease of use (least language-dependence), and also description of recommended structures for allowing language-specific variants. Liaison with national governments and telecommunications organizations, ETSI HF and the European Commission.

Relating this to terminology work on Administrative nomenclature (see clause 6.3.1).

6.1.2 eLearning (eEducation, eResearch, eWorking)

Work Item: Learning Object Model.

Deliverable/Actions:

Review (Technical Report) of the Learning Object Model CWA from the WS Learning Technologies. Assessment of localisable elements. Liaison with the WS Learning Technologies, IEEE and with ISO/IEC JTC 1 / SC 36 (*Learning technology*). Assessing common CDICT elements in eEducation, eResearch, eWorking – all related aspects of eLearning in a "lifelong learning" environment.

6.1.3 eHealth – Health online

Work Item: CDICT Domain-specific assessment for CEN / TC 251.

Deliverable/Actions:

Comparison of the many ENs, TSs and CRs and WIs in CEN / TC 251 (*Health informatics*) to CDICT work planned for CAGE to assess any common approaches, and their transferability to other domains, based on message-related standards (including XML-MAP), data standards, card applications, registration schemes, coding schemes, and terminology systems (METAVOC).

Pilot study to assess feasibility and timescale for localization (by language and use) in various types of health standard.

6.2 Work related to themes in the eEurope Action Plan for 2002

6.2.1 eAccessibility – CDICT for disabled and elderly users

Work Item: CDICT and eAccessibility – Assessment of common approaches.

Deliverable/Actions:

Review (Technical Report) covering common approaches between CDICT and eAccessibility, covering CDICT related to display, input and other issues. To cover web-based and mobile-phone-based developments in CDICT and eAccessibility.

Liaison with national governments and telecommunications organizations, ETSI HF and the European Commission.

6.2.2 eCommerce

Work Item: Standards needed for the consumer in eCommerce (project under CEN WS eCommerce, currently delayed).

Deliverable/Actions:

Adding CDICT specifications to specifications to users. Restarting delayed eCommerce WS project. Liaison with eCommerce Workshop, ISO/IEC JTC 1 / SC 32, ISO/COPOLCO.

Rationale: WS work related to users was delayed because of lack of resources. Most other eCommerce WS specifications deal with providers (system to system): this one relates to system to users. CEN eCommerce WS to oversee any CDICT work, with CAGE assisting.

6.2.3 eContent

Work Item 1: Data-elements transfer (country identifiers, language identifiers).

Deliverable/Actions:

Technical report and tables mapping MARC21 codes and ISO language identifiers.

Rationale: For Internet use, RFC 3066 (language tags) and the equivalent for country identifiers, specify the use of ISO codes (from ISO 639 and ISO 3166). Dublin Core specifications also specify the use of RFC 3066 (language tags) – or its predecessor RFC 1766. However, many of the databases which will be used to compile Dublin Core record structures use MARC21 codes, which in several cases lack a one-to-one match with ISO 639. This has not been addressed in documentation for MARC21.

Work Item 2: Extended language coding.

Deliverable/Actions:

Technical report noting developments under way in providing additional language identifiers, and noting conventions for linking language identifiers, country identifiers and date tags, to enable more specific entities to be coded. Provision of relevant information on the CAGE website.

Rationale: Work is under development, and users will need to know up to date information, which affects various different standards bodies and their users.

6.2.4 eTransport – Intelligent Transport Systems

Work Items: LOCODES in Intelligent Transport Systems.

Deliverable/Actions:

Feasibility study for the use of LOCODES (Location codes) as data elements in Intelligent Transport Systems: Assess prospects for interoperability and localizability (to account for local name variations).

Assess the openness of the LOCODES system to governmental and other use, and effectiveness of its registration system.

Rationale: The registration system seems less systematically monitored than others in the ISO system.

6.3 Other issues

Other issues identified either by the PWC project, and/or by the current Project Team, include the following:

- terminology work
- translation
- language identification
- speech applications

Most of these tend to be usable in many situations, and cross various boundaries, and underly many of the other work areas.

6.3.1 Terminology – Liaison and development

Work Item 1: Administrative nomenclature (based on work in Norway, possibly being extended within the Nordic countries).

Deliverable/Actions: XML/Topic Map descriptions of administrative structures, with translated terms and equivalences for different countries. Large project.

Rationale: With increased cross-border working in Europe, many European citizens need to deal with administrations in different countries, some of which will be subtly different in style and operation, and differences add to costs, unless they are allowed for. The deliverable would provide an informational framework for European citizens, and also provide a schematic framework that could be used in enhancing eGovernment development.

Work Item 2: Input from Europe into a new ISO / TC 37 preliminary work item on multilingual product classification, “Basic principles of multilingual product classification for electronic commerce.”

Deliverable/Actions: Structured mapping of terms in various European languages (also taking into account country differences within languages) to standard product classification codes. Liaison with eCommerce WS, ISO / TC 37 and ISO/IEC JTC 1 / SC 32.

Work Item 3: Domain-related terminology (eHealth, eTransport).

Deliverable/Actions: Development of terminology data for those domains. Liaison with ISO / TC 37 over terminology and with ISO/IEC JTC 1 / SC 32 on topic maps and other specifications.

Rationale: Both eHealth-related terminology and eTransport-related terminology work items are likely to arise from the eHealth and eTransport areas (see clauses 6.1.3 and 6.2.4). Other domains may also benefit. Consistent approaches will be necessary for such work.

6.3.2 Translation – Liaison and development

Work Item: Technology watch and liaison.

Deliverable/Actions: Liaison, technology watch, dissemination (via website). Liaison with the European Translators Platform, LISA, ETSI and other bodies. Report(s) on best practice.

Rationale: This area has not been covered previously in CEN, but it is likely that some standardization issues will arise as machine translation (MT) is increasingly routinely offered on the web, and other devices such as automated cash dispensers in banks, and via telephone services.

Ways to identify and offset specific limitations of MT (strictly only ever likely to be machine-assisted translation except in very limited domains) also need to be identified. Ongoing project.

6.3.3 Language identification

Work item: Language identifiers: present situation and future development.

Deliverable/Actions: Technical report on provision on the Internet for European languages with no language identifiers. Liaison with ISO / TC 37 / SC 2, IETF, IANA, SIL, LREC and others. Technology watch, and dissemination (via website).

Rationale: Language identifiers are routinely in various XML transactions (especially involving xml:lang) and in bibliographic information interchange, and in various other situations on the Internet, including in MIME in email and for various web pages.

Language coding is typically based on ISO 639-1 and ISO 639-2, though other coding systems exist such as SIL and FIPS language codes. Language tagging for specific implementations of these such as MARC21 and RFC 3066 define specific implementations of ISO 639 identifiers.

CEN / TC 304 reported a lack of identifiers for various European languages, five years ago, and there are still no identifiers for several that were reported then. This is compounded by the use of different coding methods for different languages in different situations.

New initiatives are reported in ISO / TC 37 / SC 2 and also in other arenas: European users urgently need up to date information on the current status and likely future developments.

6.3.4 Speech applications – Liaison and development

Work Item: Technology watch and liaison.

Deliverable/Actions: Liaison, technology watch, dissemination (via website). Liaison with W3C, ETSI and other bodies. Report(s) covering best practice, and implications for use of specific languages (“standard” forms and dialects).

Rationale: This area has not been covered previously in CEN, but it is likely that some standardization issues will arise as technology changes, and as takeup increases.

Speech input and speech output is used in public address systems attached to messaging systems and databases, and also Interactive Voice Response (IVR) in telephone transactions, and in dictation on computers.

Speech interaction, including speech recognition and speech output, has been routinely used in certain well-specified areas of telephone communications, and other applications are emerging for speech recognition, input and output in computer systems.

It has many actual or potential uses, including in computer-assisted language learning (CALL), email reading to visually/physically impaired people, use in call centres, and many others.

In web development, W3C's CSS2 specification (1998) allows for the generation of speech from text on the web.

Speech-enablement of other devices, e.g. in some cars, game, mobile and/or multimedia devices, also occurs, though in proprietary rather than standard ways, as that market sector is still very new. Technology changes rapidly, and so do companies involved in delivering some of these products, so it is important to be able to distinguish transient practices from emerging standards.

Links to other work areas in the work programme

Standardization for speech applications will have many links with most other work areas and work items listed in the work programme. Some examples are listed below.

There are obvious applications in eAccessibility, eCommerce, eTransport, where speech input or output can assist users.

Text to/from speech can also be linked to translation between two specific languages. There are also some analogies with translation in the reliance on specific languages, and on constant source-to-target conversion.

There is a reliance on terminology management, for terms, phrases and large messages. Significant work has been done for some major languages in Europe: additional work will need to be done for other languages in Europe.

There is a need to add language tags to speech data, e.g. for languages outside the current scope of in some cases beyond current distinctions between standard versions of Netherlands Dutch and Belgian Dutch, Nynorsk and Bokmål in Norway, and between English, French, Spanish and Portuguese accents in Europe and their counterparts in the Americas.

6.4 Progression of previous work under BC/CEN/97/26

The Project Team is also asked to recommend how the outstanding items under BC/CEN/97/26 should be progressed (if at all). In summary, the Project Team endorses the CEN/BT Resolution BT 12/2002 on making CEN / TC 304 dormant, and suggests a further look at some of the previous work. The detail is given below.

The Project Team endorses Resolution BT 12/2002 (14th TCMG item: 2.3 and 2.6), concerning the "Status and Secretariat of CEN / TC 304" from the "Resolutions taken at the 14th BT/TCMG meeting Brussels, 2002-06-25." This stated (summarised, and in a slightly different order):

(a) There is potentially important work for Europe on Cultural Diversity in ICT systems, currently overseen by the CEN/ISSS Cultural Diversity Steering Group.

(b) Uncompleted work should be withdrawn or cancelled.

(c) SIS should take on the CEN / TC 304 Secretariat, just to progress the remaining work items (00304030, 00304051 and 00304052) to adopted CEN publication status, and that CEN / TC 304 should become dormant after that.

(d) The CEN/ISSS Cultural Diversity Steering Group (CDSG) should be asked to evaluate the future direction for the published standards EN 1922:1997 (*Character set … for interworking with Telex services*) and ENV 13710:2000 (*European Ordering Rules – Ordering of characters from the Latin, Greek and Cyrillic scripts*).

We also recommend that CAGE look briefly at some of the issues below, where any loose ends are highlighted, though this tackling of loose ends should have a lower priority than the main developments.

As it stands, the only definite ongoing work would be the three following work items, and their remaining life cycle is very short, as they are near their completion stages:

1. European Fallback Rules (ch. to TR) 26.10.2
2. European extension to ISO repertoire of OCR-B (EN) (built on Test results) 26.15
3. Revision work on EN 1923 (Inclusion of euro sign and missing European letters in new TS 1923)

Those would complete earlier aspects of the work of CEN / TC 304.

The following work would be deleted if CEN / TC 304 is made dormant, overriding resolutions of CEN / TC 304:

4. CWA Eurolocale (was due to become a TS)
5. Population of Cultural Registry: guide (26.19)
6. Support for Sami languages (EN) (CEN / TC 304 Resolution M19#10)
7. EOR-1 and -2 merged into an EN (CEN / TC 304 Resolution M19#2)

We also comment briefly on some of those below, as some intermediate work has been done.

The Project Team recommends that items 1, 2 and 3 above (work items 00304030, 00304051 and 00304052) remain within CEN / TC 304 prior to dormancy of CEN / TC 304. If necessary, CAGE could pursue some aspects later, if they suddenly regained importance.

Their life cycle will be finished soon in any case, due to the natural course of balloting and dealing with editorial comments, and no additional costs are involved. These are also discussed further below, as some of these, and some other completed work items noted below, may need to be looked at later, for reasons of consistency with other standards, though with a lesser priority.

6.4.1 European Fallback Rules (WI 00304030/prCR 14381)

For WI 00304030/prCR 14381 (*European Fallback Rules – Fallbacks for Latin, Greek and Cyrillic scripts*), this is still in the state of waiting for approval by BT under draft Resolution BT C27/2002.

Two countries indicated fundamental disagreement with the work during the last vote according to recent advice by the CEN / TC 304 Chair, via the CEN / TC 304 email list (TC304.2507).

The Project Team therefore recommends that this receive a final examination to highlight any problem areas before it is issued. There is another reason for CAGE to review it, as there are also other fallback/transliteration conventions that are in widespread use, which are not covered in the European Fallback Rules (WI 00304030/prCR 14381), outlined below.

6.4.1.1 Fallback, UNGEGN and ICU

There may be a technical issue relating to whether some of the Fallback conventions described may any longer relate to industry practice. The widely used ICU classes maintained by IBM increasingly use the fallback and transliteration conventions from UNGEGN (the United Nations Group of Experts on Geographic Names) which are used on a worldwide basis in maps and gazetteers, and which through ICU specifications are likely to be used more widely.

There should be a brief technical review of the current prCR 14381 to assess whether any of its recommendations are in fact now outdated, in relation to any of the fallbacks for particular scripts.

In August 2002 the Eighth United Nations Conference on the Standardization of Geographical Names will meet in Berlin, as will its Romanization Working Group. We recommend that some CDSG representation should be organized.

Peeter Päll (Estonia) is now the convenor of the UNGEGN Romanization Working Group, and himself involved in Estonia with other aspects of CDICT.

We recommend that CAGE establishes liaison with the UNGEGN Romanization Working Group.

6.4.1.2 Fallback, ICAO 9303 and ISO/IEC 7501 (Machine-Readable Passports)

In addition, given the other major user of fallback in actual systems (viz. in passport-related applications, which is overseen by ISO/IEC JTC 1 / SC 17), CDSG in the interim should also contact Michael Ellis (AU), Project Leader of ISO/IEC JTC 1 / SC 17 / WG 3 / Task Force 3 (Transliteration).

ISO/IEC 7501 (*Machine-readable passports*) is in process of revision, and ICAO 9303 (*Machine-readable travel and related ICAO documents*) has recently been revised.

ICAO 9303 also contains fallback/transliteration tables that will be in very widespread use, given their use in passports, and revision of ISO/IEC 7501 has always taken its lead from previous editions of ICAO 9303 in the past.

There are also implications in these for use of fallback in Schengen countries, and in Europe as a whole: problems relating to this have resulted in legal cases, in national supreme courts and in the European courts.

This is not just "another TR": This CEN TR does need to be looked at in advance prior to it being issued, and if necessary revised before issue, and/or guidance issued, given the complications and costs which sometimes arise for European governments, in the area of names in passports.

CEN / TC 304 originally had a work item for names in passports, but only OCR-B issues were dealt with in this in the end.

6.4.2 WI 00304051 (European addition to OCR-B)

A draft EN on the European addition to OCRB locale (work item 00304051) has been sent to the CEN Management Centre for submission to CEN Enquiry.

We also recommend that it should also be investigated how far this relates to work on Fallback (see discussion in clause 6.4.1) and similar issues in ICAO and in ISO/IEC JTC 1 / SC 17 and in industry practice, and whether there are any differences in approach, and what cost implications are involved if there are any differences in practice.

Liaison with ISO/IEC JTC 1 / SC 17 / WG 3 / Task Force 2 (*Standards Harmonization*, Project Leader: Charles Chatwin, UK) would also be useful in this regard.

6.4.3 WI 00304052/prCEN/TS 1923 rev (European Character Repertoires: 8-bit single byte coding)

A CEN TS was developed in CEN / TC 304 (work item 00304052, to become the future TS 1923). This was submitted for vote with a deadline of 2002-04-11. No additional comments are made here.

6.4.4 Eurolocale

This was resolved (CEN / TC 304 M19#6) by arranging for communication with ISO/IEC JTC 1 / SC 22 / WG 20 on this issue. No additional comments are made here.

6.4.5 Population of Cultural Registry: guide (26.19)

ENV 12005:1996 has been transformed into an ISO/IEC standard, ISO/IEC 15897 (as its use potentially covers far more than just European needs). CEN BT therefore resolved that the original European standard, ENV 12005:1996 (*Information technology – Procedures for European registration of cultural elements*) should be withdrawn by CEN (see Resolution BT 12/2002 from the "Resolutions taken at the 14th BT/TCMG meeting Brussels, 2002-06-25").

The only outstanding work item relating to this was the preparation of a guide. Resolution CEN / TC 304 M19#5 notes that document CEN / TC 304 N1002 Monica Ståhl Forsberg contains guidance to help European NBs prepare cultural information for registration.

That effectively closes this work item anyway, and the register is in any case now maintained on behalf of ISO/IEC JTC 1 / SC 22 / WG 20, and no longer for CEN / TC 304.

However, the Project Team recommends that in due course, CAGE looks at the following anomalies, and liaise accordingly with ISO/IEC JTC 1 / SC 22 / WG 20:

(a) There are very few actual registry entries for cultural conventions. There are however, a large number of charset entries, which give a misleading view of the number of registrations).

(b) The syntax used in register entries ignores the conventions used in ISO/IEC 10646 and Unicode for recording character identifiers (U+1234 etc) and uses others which are not widely used.

(c) For charsets, there are also other mechanisms in other bodies, such as with some IANA and IETF mechanisms, for recording similar data, that may be of more use.

We recommend that CAGE liaise with ISO/IEC JTC 1 / SC 22 / WG 20 regarding whether the syntax for recording registrations should be amended or whether the register needs to exist in its present form, given the lack of registrations. Comments from CAGE might seek either

(i) to stimulate the preparation of more register entries, or

(ii) to suggest other mechanisms other than the present register, if these ways of recording cultural conventions turn out not to be widely used.

6.4.6 Support for Sami (EN)

As work in parallel on this has been started in the Nordic countries Finland, Norway and Sweden to produce a document supported by the main interested parties in these countries for Sami languages, no additional work needs to be done by CAGE, though it would be useful for the CDICT website to include information on developments here.

6.4.7 Unifying two parts of European Ordering Rules (ENV 13710)

CEN / TC 304 Resolution M19#2 resolved that the two parts of European Ordering Rules (ENV 13710), EOR-1 and EOR-2, should be merged into an EN. Project leaders were allocated to that.

In Resolution BT 12/2002, the CEN/ISSS Cultural Diversity Steering Group (CDSG) was asked to evaluate the future direction for ENV 13710:2000 (*European ordering rules – Ordering of characters from the Latin, Greek and Cyrillic scripts*).

The Project Team recommends that CAGE takes on responsibility for this, rather than CDSG. However, it should not be bound by earlier decisions of CEN / TC 304, but if necessary should take a fresh look at ENV 13710:2000 (EOR-1), at EOR-2, and the relationship of each to the underlying base standard ISO/IEC 14651.

6.4.8 Additional standards

Resolution BT 12/2002 also asked the CEN/ISSS Cultural Diversity Steering Group (CDSG) to evaluate the future direction for the published standard EN 1922:1997 (*Character set ... for interworking with Telex services*).

CAGE should look at this and other older CEN standards again, and in view of the Project Team's recommendations about not reinventing base standards that were generally developed maintained by ISO/IEC JTC 1 / SC 2 and other ISO/IEC JTC 1 committees, to consider withdrawing it.

7 Draft Business Plan

Note: This chapter is a draft business plan for CAGE (CDICT Advisory Group for Europe), the body that is proposed to carry out the work programme outlined in the previous section. It depends on its approval by CEN, and on any changes made during that process.

The rationale for CAGE derives from a series of reports to CEN, and recommendations by high-level committees within CEN:

- the PWC report on standardization/consensus-building measures, and recommendations on CEN / TC 304 (*Information and communications technologies – European localization requirements*);
- Resolution CEN BT 12/2002 (June 2002), concerning the "Status and Secretariat of CEN / TC 304" and the CEN / TC 304 reply;
- recommendations from the CEN / TC 304 Chair (June 2002) about possible future status of existing work in a more dormant CEN / TC 304.
- the report of the Project Team on Programming for Cultural Diversity Consensus-Building, and subsequent decisions by CEN.

7.1 Title: *CDICT Advisory Group for Europe* (CAGE)

This title is provisional. It is used as a working title and acronym for the purpose of reference in this report.

7.2 Scope

CAGE should advise on coordination of activities at the technical level, for issues relating to CDICT (Cultural Diversity in ICT systems), including recommending allocation and supervision of work, either within CAGE, elsewhere within CEN, or outside of CEN. It should take account of the market environment, economic factors, social factors and technical factors in its advice.

Liaison, technology watch, and dissemination should all be a major part of its focus.

(a) Liaison should be active, with a range of bodies where CDICT is important, and which are likely to depend on input from CEN.

(b) Technology watch should result in the identification of new areas, as they arise, where CDICT may be an issue. It should also result in identification of superseded technologies and/or standards where appropriate, given the rapid life cycle of new technologies in ICT.

(c) Dissemination should include the operation of a website, and also one-off workshops/conferences, in conjunction with other bodies, on CDICT issues in Europe.

CAGE should operate as a service to CEN, its TCs and its WSs, and liaise with those and with other bodies concerned with standards which involved CDICT, in particular with those TCs and WSs of CEN which are active in developing standards and/or specifications in the following areas which are related to the 2005 eEurope Action Plan:

- eGovernment,
- eLearning, and
- eHealth

and the 2002 eEurope Action Plan:

- eAccessibility,
- eCommerce,
- eContent, and
- eTransport

and in addition cover the following aspects:

- terminology work,
- translation,
- language identification,

- speech applications.

Finally, there remains a very small amount of work remaining from BC/CEN/97/26, which results from previous work by CEN / TC 304 prior to it becoming dormant.

It should not cover the following areas, which were identified in the PWC report except in exceptional cases, as either

(a) others take a lead here, and/or

(b) most practice already follow certain standards, so additional work is not necessary, or in some cases

(c) CDICT aspects make little impact on some of these areas: Telecommunications infrastructures and connection policies, network protocols, data formats and metadata, electronic commerce, mark-up language internationalization, character sets, alphabetical ordering, character matching, keyboards.

Aspects like writing style should not be covered, except as information in the Cultural Registry, if appropriate.

Nor should it duplicate existing work in ISO or industry consortia, nor just make European profiles of these, as such information is often readily available in any case. In particular, the principal published standards of the CLAUI TD (ISO/IEC 10646, ISO/IEC 14651, and ISO/IEC 9995) do not need European profiles.

It may however add any Europe-specific aspects that may be unavailable in a more international specification, and also initiate any additional new work at a global level, if it is clear that Europe has needs in this area, but that no other bodies are yet taking a lead. However, if such bodies emerge later, it should be happy to pass over such global level work to those committees.

It should aim at developing specifications that can be adopted in various domains, and may well need a new focus, e.g. concentrating on specifications that can be used in XML implementations.

7.3 Modus operandi

The CDICT Advisory Group for Europe (CAGE) should operate as an advisory group drawn from standards, government and industry experts within Europe, to provide guidance on how European needs for standardization in the area of cultural diversity can best be met. It should be able to interact with existing standards committees, workshops and consortia, in ways that allow those other bodies to be able to provide CDICT provisions within their own standards and specifications.

In particular, outside of CEN, its members should pay special attention to those bodies involved in the Cultural and Linguistic Adaptability and User Interface Technical Direction (CLAUI TD) of ISO/IEC JTC 1 (i.e. ISO/IEC JTC 1 / SC 2 / WG 2, ISO/IEC JTC 1 / SC 22 / WG 20 and ISO/IEC JTC 1 / SC 35).

Recommending lead responsibilities for proposed new work items should be part of its brief. Where necessary, CAGE members should be in a position where they are able to provide expertise to carry out some technical work, but more often they should be in a position to suggest other fora and groups where such work is going on, or likely to go on, and to ensure that developments there meet European needs, and also to avoid concentrating too much load on too few experts.

CAGE should also be in a position to recommend funding sources for any work that needs to be carried on within CAGE, either from funding by the European Commission, from national governments, or through sponsorship by industry, either directly through funding or indirectly through provision of expert time and/or services.

CAGE should meet twice a year, on the same basis as CDSG. Where necessary, and by prior agreement, decisions could be made by correspondence (email and telephone) with records as necessary.

Participants should initially be selected on the same basis as the representatives selected for the Cultural Diversity Steering Group (CDSG), though widening this may be useful at a later stage.

Representation of national member bodies, per se, is not envisaged. If (exceptionally) work items arise in the future which specifically recommend only national member body input, the feasibility of developing such work items within CEN / TC 304 should be investigated.

Occasionally there may also be a need for the more formal relationships with ISO under the Vienna Agreement, which the existence of CEN / TC 304 may facilitate.

7.4 Dissemination

With regard to dissemination, the CAGE website/portal should arrange to work with, and to be linked to, other portals that monitor CDICT standards to any extent, e.g. the Diffuse project pages and the Cover pages.

The European Union's IDA programme supports interoperability, standardisation and the provision of pan-European services, and the IDA portal now being developed plans to offer multilingual access points for online information and services to citizens and enterprises supporting mobility across borders.

It is also planned that the IDA portal should serve as test bed for multi-platform technologies, including the use and improvement of machine translation systems to fully cover the linguistic diversity of European users of such web sites.

We recommend that CAGE cooperate with the IDA and DIFFUSE projects and portals, and with any successors to these.

The CAGE portal should also add information about status and takeup, and/or recommended deprecation, of particular standards or specifications, much as some of the Diffuse pages already do.

Where possible, CAGE members should also get involved with EU Framework Plan proposals and evaluation, and various similar activities, to ensure that CDICT standards are not taking place in a vacuum, and that appropriate standards and specifications are taken up in such Framework Plan proposals and projects.

They should also attempt to be involved with Chambers of Commerce, ICT trade associations, and similar bodies in their own countries.

7.5 Funding

Adequate funding for these various activities is vital for their success.

We recommend that at the earliest opportunity, CAGE prepares a funding proposal for the technical work such as that outlined in the previous section, for the various work items, particularly those which relate to ongoing or future work in the eEurope Action Plan. This should be done in conjunction with CEN/ISSS and with other TCs and/or WSs of CEN, where appropriate, so that it can be made part of CEN's other requests for funding from the European Commission.

In addition, for some aspects, funding from industry may be appropriate, especially where industry can make a sponsorship arrangement. Examples of such funding may be particularly appropriate in dissemination activities, where a company may wish to indicate its association with CDICT activities and with CAGE by a prominent display of its name and logo, either on the web, or in association with a workshop/conference or exhibition.

Such a mix of funding would underline the joint participation of government and industry in CDICT in Europe.

8 References

This report is available in various formats at http://www.hjulstad.com/havard/CDIV/.

Sites that have been referenced in the text:

- ANEC – European Association for the Co-ordination of Consumer Representation in Standardisation – http://www.anec.org/.
- CEN WS/eBES – e-Business Board for European Standardization – http://www.cenorm.be/isss/Workshop/eBES/.
- CEN WS/EC – Electronic commerce – http://www.cenorm.be/isss/Workshop/ec/.
- CEN WS/KM – Knowledge management – http://www.cenorm.be/isss/Workshop/km/.
- CEN WS/LT – Learning technologies – http://www.cenorm.be/isss/Workshop/LT/.
- Diffuse – http://www.diffuse.org/.
- ECMA – Standardizing Information and Communication Systems – http://www.ecma.ch/.
- ETSI – European Telecommunications Standards Institute – http://www.etsi.org/.
- IANA – Internet Assigned Names Authority – http://www.iana.org/.
- ICAO – International Civil Aviation Organization – http://www.icao.int/.
- IETF – Internet Engineering Task Force – http://www.ietf.org/.
- ISO – International Organization for Standardization – http://www.iso.org/.
- ISO 639 Registration Authorities' Joint Advisory Committee – http://lcweb.loc.gov/standards/iso639-2/iso639jac.html.
- ISO 3166 Maintenance Agency – http://www.iso.org/iso/en/prods-services/iso3166ma/.
- ISO / TC 37 – http://linux.infoterm.org/iso-e/i-iso.htm.
- ISO/IEC JTC 1 – http://www.jtc1.org/.
- ISO/IEC JTC 1 CLAUI – http://std.dkuug.dk/jtc1/claui/.
- ISO/IEC JTC 1 / SC 2 – http://std.dkuug.dk/jtc1/sc2/.
- ISO/IEC JTC 1 / SC 17 – http://www.sc17.com/.
- ISO/IEC JTC 1 / SC 22 / WG 20 – http://std.dkuug.dk/JTC1/SC22/WG20/.
- ISO/IEC JTC 1 / SC 32 – http://www.jtc1sc32.org/.
- ISO/IEC JTC 1 / SC 35 – http://forum.afnor.fr/afnor/WORK/AFNOR/GPN2/Z62A/index.htm.
- ISO/IEC JTC 1 / SC 36 – http://jtc1sc36.org/.
- ITU – International Telecommunication Union – http://www.itu.int/.
- ITU-T, the standardization study group – http://www.itu.int/ITU-T/.
- LISA – Localisation Industry Standard Association – http://www.lisa.org/.
- PWC – PricewaterhouseCoopers – http://www.pwcglobal.com/.
- SIL International – http://www.sil.org/.
- Unicode Consortium – http://www.unicode.org/.
- W3C – World Wide Web Consortium – http://www.w3.org/.

Annex A – Languages in Europe

The following is a listing by country of most of the "indigenous" languages of Europe, including languages with a formal "status" ("national language", "regional language", etc.).

Counting languages is impossible without very clear definitions of what an "individual language" actually is. The language names given here are derived from several sources, each of which may have a different view on this issue. The purpose of the different source documents also differs.

The procedures of assigning "status" to certain languages also vary according to local legislation and other factors. Even in this process the counting of languages is an issue in various ways. One example is "the Sami language", which has "official status" in Finland, Norway, and Sweden. However, there exists no single "Sami language". Most linguists tend to assume that at least five Sami languages must have been granted "status" in the three countries combined. A similar case is the Frisian languages, which have been treated differently in the "status assignment" in Germany and the Netherlands. There are a number of other cases as well, including Romany, Ruthenian, and Albanian.

For each country is indicated CEN membership ("M" = CEN member, "A" = CEN affiliate, "C" = CEN correspondent). By "Charter" is indicated that the country is a signatory to the European Charter for Regional or Minority Languages (as of 2002-06-25).

CEN members (M) are the national standards bodies of the EU and EFTA countries, the Czech Republic and Malta. **CEN affiliates (A)** are: Albania, Bulgaria, Croatia, Cyprus, Estonia, Hungary, Latvia, Lithuania, Poland, Romania, Slovakia, Slovenia, and Turkey. **CEN correspondents (C)** are: Ukraine and Yugoslavia, and the non-European countries Egypt and South Africa.

Note: "Main languages" are normally official, national languages. "Languages with status" are additional languages with some sort of official status, in most cases as recorded in declarations in connection with the European Charter for Regional or Minority Languages. "Other languages" lists other languages based on various sources. The list is by no means complete.

For each country there is a "diversity index". That figure is taken from publications by SIL International. It indicates the probability that two random persons within the country will have different first languages. The calculation has the problem that relates to the counting of languages. E.g. "Serbo-Croatian" is probably counted as one language in this calculation. The calculation probably does not include immigrant languages. The Project Team has no opinion about the value of this "diversity index".

Albania (A) – *Main languages:* Albanian – *Other languages:* Greek, Macedonian – *Diversity index:* 0.26
Andorra – *Main languages:* Catalan – *Other languages:* Spanish – *Diversity index:* 0.57
Armenia (Charter) – *Main languages:* Armenian – *Languages with status:* Assyrian, Yezidi, Greek, Russian, Kurdish – *Other languages:* Azerbaijani, Kurmanji – *Diversity index:* 0.16
Austria (M) (Charter) – *Main languages:* German – *Languages with status:* Burgenlandcroatian, Slovenian, Hungarian, Czech, Slovak, Romany – *Diversity index:* 0.14
Azerbaijan (Charter) – *Main languages:* Azerbaijani – *Other languages:* Avar, Kurmanji, Lezghian, Talysh – *Diversity index:* 0.37
Belarus – *Main languages:* Belarusian – *Diversity index:* 0.36
Belgium (M) – *Main languages:* Dutch, French, German – *Other languages:* Luxembourgish – *Diversity index:* 0.65
Bosnia and Herzegovina – *Main languages:* Bosnian, Croatian, Serbian – *Diversity index:* 0.00
Bulgaria (A) – *Main languages:* Bulgarian – *Other languages:* Albanian, Gagauz, Romany, Turkish – *Diversity index:* 0.22
Croatia (A) (Charter) – *Main languages:* Croatian – *Languages with status:* Italian, Serbian, Hungarian, Czech, Slovak, Ruthenian, Ukrainian – *Diversity index:* 0.07
Cyprus (A) (Charter) – *Main languages:* Greek, Turkish – *Diversity index:* 0.37
Czech Republic (M) (Charter) – *Main languages:* Czech – *Other languages:* German, Polish, Romany – *Diversity index:* 0.06
Denmark (M) (Charter) – *Main languages:* Danish, Faroese, Greenlandic – *Languages with status:* German – *Diversity index:* 0.05
Estonia (A) – *Main languages:* Estonian – *Diversity index:* 0.48
Finland (M) (Charter) – *Main languages:* Finnish, Swedish – *Languages with status:* Sami, Romany – *Diversity index:* 0.14
France (M) (Charter) – *Main languages:* French – *Other languages:* Alemannisch, Basque, Breton, Catalan, Corsican, Dutch, Franco-Provençal, Gascon, Italian, Provençal – *Diversity index:* 0.24

Georgia – *Main languages:* Georgian – *Other languages:* Abkhazian, Mingrelian, Osetin – *Diversity index:* 0.57
Germany (M) (Charter) – *Main languages:* German – *Languages with status:* Danish, Upper Sorbian, Lower Sorbian, North Frisian, Sater Frisian, Romany, Low German – *Other languages:* Polish, Yiddish – *Diversity index:* 0.18
Gibraltar – *Main languages:* English – *Other languages:* Spanish – *Diversity index:* 0.50
Greece (M) – *Main languages:* Greek – *Other languages:* Albanian, Bulgarian, Romanian, Turkish – *Diversity index:* 0.14
Hungary (A) (Charter) – *Main languages:* Hungarian – *Languages with status:* Croatian, German, Romanian, Serbian, Slovak, Slovenian – *Other languages:* Romany – *Diversity index:* 0.14
Iceland (M) (Charter) – *Main languages:* Icelandic – *Diversity index:* 0.00
Ireland (M) – *Main languages:* Irish, English – *Diversity index:* 0.17
Italy (M) (Charter) – *Main languages:* Italian – *Other languages:* Albanian, Franco-Provençal, French, Friulian, German, Greek, Ladin, Ligurian, Lombard, Neapolitan, Provençal, Sardinian, Sicilian, Slovenian – *Diversity index:* 0.59
Latvia (A) – *Main languages:* Latvian – *Other languages:* Yiddish – *Diversity index:* 0.60
Liechtenstein (Charter) – *Main languages:* German – *Diversity index:* 0.13
Lithuania (A) – *Main languages:* Lithuanian – *Diversity index:* 0.34
Luxembourg (M) (Charter) – *Main languages:* Luxembourgish, French – *Other languages:* German – *Diversity index:* 0.53
Macedonia (former Yugoslav Republic) (Charter) – *Main languages:* Macedonian – *Other languages:* Albanian, Romanian, Romany, Turkish – *Diversity index:* 0.49
Malta (M) (Charter) – *Main languages:* Maltese, English – *Diversity index:* 0.02
Moldova – *Main languages:* Moldovan – *Other languages:* Bulgarian, Gagauz, Romany – *Diversity index:* 0.59
Monaco – *Main languages:* French – *Other languages:* Ligurian, Provençal – *Diversity index:* 0.52
Netherlands (M) (Charter) – *Main languages:* Dutch – *Languages with status:* Frisian, Limburgish – *Diversity index:* 0.20
Norway (M) (Charter) – *Main languages:* Norwegian Bokmål, Norwegian Nynorsk – *Languages with status:* Sami – *Other languages:* Finnish – *Diversity index:* 0.08
Poland (A) – *Main languages:* Polish – *Other languages:* Belarusian, German, Kashubian, Romany, Ukrainian – *Diversity index:* 0.12
Portugal (M) – *Main languages:* Portuguese – *Other languages:* Galician – *Diversity index:* 0.02
Romania (A) (Charter) – *Main languages:* Romanian – *Other languages:* Bulgarian, German, Hungarian, Romany, Turkish – *Diversity index:* 0.20
Russia (Charter) [European part] – *Main languages:* Russian – *Other languages:* Adyghe, Avar, Bashkir, Chechen, Chuvash, Dargwa, Erzya, Ingush, Kabardian, Kalmyk, Karachay, Karelian, Komi, Kumyk, Lak, Lezghian, Mari, Moksha, Tatar, Udmurt – *Diversity index:* 0.27
San Marino – *Main languages:* Italian – *Diversity index:* 0.00
Slovakia (A) (Charter) – *Main languages:* Slovak – *Languages with status:* Bulgarian, Croatian, Czech, German, Hungarian, Polish, Romany, Ruthenian, Ukrainian – *Diversity index:* 0.25
Slovenia (A) (Charter) – *Main languages:* Slovenian – *Languages with status:* Romany – *Other languages:* Hungarian, Italian – *Diversity index:* 0.17
Spain (M) (Charter) – *Main languages:* Spanish – *Languages with status:* Basque, Catalan, Balearian, Galician, Valencian, Navarra – *Other languages:* Asturian – *Diversity index:* 0.44
Sweden (M) (Charter) – *Main languages:* Swedish – *Languages with status:* Sami, Finnish, Meänkieli (Tornedal Finnish), Romany, Yiddish – *Diversity index:* 0.36
Switzerland (M) (Charter) – *Main languages:* French, German, Italian, Rhaeto-Romance – *Other languages:* Lombard – *Diversity index:* 0.53
Turkey (A) [European part] – *Main languages:* Turkish – *Other languages:* Armenian, Gagauz, Bulgarian – *Diversity index:* 0.25
Ukraine (C) (Charter) – *Main languages:* Ukrainian – *Other languages:* Hungarian, Jakati, Romanian – *Diversity index:* 0.48
United Kingdom (M) (Charter) – *Main languages:* English – *Languages with status:* Welsh, Gaelic (Scottish), Irish – *Other languages:* Lowlands Scots – *Diversity index:* 0.07
Vatican State – *Main languages:* Latin – *Other languages:* Italian – *Diversity index:* 0.00
Yugoslavia (C) – *Main languages:* Serbian – *Other languages:* Albanian, Croatian, Hungarian, Romanian, Romany, Slovak – *Diversity index:* 0.32

Summary counts

To get a rough (conservative) estimate of the number of languages in Europe, the figures below have been derived from the tables above in this document. “CEN member countries” here include affiliate and correspondent members.

- Number of “main languages” in all European countries: 44
- in CEN member countries: 34
- Number of “languages with status” (including “main”) in all countries: 67
- in CEN member countries: 55
- Total number of languages in all European countries: 109
- in CEN member countries: 76

Annex B – Deliverables from CEN / TC 304

This annex relates to clause 2.1.1.1 in this report.

Former standards (withdrawn)

- **CEN ENV 4150x**-standards:
  - 8-bit (Latin 1) coded character set
  - Teletex character set and coding
  - Videotex character set and coding
  - Telex character set and coding
  - Line coding character set and coding
  - Eastern European character set and coding
- **CEN ENV 1260** – Cyrillic character set and coding
- **CEN ENV 1973** – 16 bit MES and EES
- **CEN ENV 12005** – *European Cultural Registry*

Current standards

- **CEN EN 1922** – Telex character set and coding
- **CEN EN 1923** – 8-bit coded character sets
- **CEN ENV 13710** – *European Ordering Rules-1*

Reports

- **CEN CR 13907:2000** – *General Model for Graphic Character Transformations*
- **CEN CR 13928:2000** – *Guide for Users*
- **CEN CR 14270:2001** – *European Keyboards*
- **CEN CR 14400:2001** – *European Ordering Rules-2*
- **Internal report** – *Testing of euro sign + European letters for OCR-B* (2001)
- **Report** – *The Euro in IT-standardization* (Report to CEC 1999-04-23)
- **Internal study report** – *Matching-1* (2001)
- **Internal report + Advisory report** – *Population of Cultural Registry-1 with National Locales* (2001)

Changed/cancelled work items

- **CEN TAP/LE-study** – *R&D output/experiences* (changed to PWC study)**CEN CR/CWA** – *ERR European regulatory requirements, phase 1 + 2* (cancelled due to overlap with other TCs)
- **CEN prCWA** – *ALPHA Character Repertoires of Europe* (cancelled due to lack of consensus)

Halted work items for CDSG decision

- **CEN TR/CWA** – *ESR-2 European culturally specific requirements*
- **CEN** – *Taxonomy/mapping activities*
- **CEN TR/TS** – *Matching-2*

Work items (proposed) deleted

- **CEN TS** – *Eurolocale* (Transform CWA Eurolocale part 1 to TS)

- **REGISTRATION** – *Population of Cultural Registry-2*
- **CEN EN** – *Support for Sami languages*
- **CEN EN** – *EOR-1 and -2 merged*

Ongoing work items (Secretariat: SIS)

- **CEN prTR 14381** (BT 6589) – *European Fallback Rules*
- **CEN prTS 1923** – *Revision work on EN 1923* (Inclusion of euro sign and missing European letters)
- **CEN EN** – *European extension to ISO repertoire of OCR-B*

CWAs (related to CEN / TC 304)

- **CEN CWA 13873:2000** – *Multilingual European Subsets in ISO/IEC 10646*
- **CEN CWA 14051-1:2000** – *European generic locales (Eurolocale) – Part 1: General Specifications*
- **CEN CWA 14051-2:2000** – *European generic locales (Eurolocale) – Part 2: Narrative cultural specifications, POSIX locales, and repertoire map*
- **CEN CWA 14094:2001** – *European Culturally Specific Requirements (ESR-1)*

Annex C – Some standards from ISO/IEC JTC 1 / SC 2

This list of International Standards relates to clause 2.2.3 of this report.

- **ISO/IEC 646:1991** – *Information technology – ISO 7-bit coded character set for information interchange.*
- **ISO 1073-1:1976** – *Alphanumeric character sets for optical recognition – Part 1: Character set OCR-A – Shapes and dimensions of the printed image.*
- **ISO 1073-2:1976** – *Alphanumeric character sets for optical recognition – Part 2: Character set OCR-B – Shapes and dimensions of the printed image.*
- **ISO 5426:1983** – *Extension of the Latin alphabet coded character set for bibliographic information interchange.*
- **ISO 5427:1984** – *Extension of the Cyrillic alphabet coded character set for bibliographic information interchange.*
- **ISO 5428:1984** – *Greek alphabet coded character set for bibliographic information interchange.*
- **ISO 6936:1988** – *Information processing – Conversion between the two coded character sets of ISO 646 and ISO 6937-2 and the CCITT international telegraph alphabet No. 2 (ITA 2).*
- **ISO/IEC 6937:2001** – *Information technology – Coded graphic character set for text communication – Latin alphabet.*
- **ISO/IEC 8859-1:1998** – *Information technology – 8-bit single-byte coded graphic character sets – Part 1: Latin alphabet No. 1.*
- **ISO/IEC 8859-2:1999** – *Information technology – 8-bit single-byte coded graphic character sets – Part 2: Latin alphabet No. 2.*
- **ISO/IEC 8859-3:1999** – *Information technology – 8-bit single-byte coded graphic character sets – Part 3: Latin alphabet No. 3.*
- **ISO/IEC 8859-4:1998** – *Information technology – 8-bit single-byte coded graphic character sets – Part 4: Latin alphabet No. 4.*
- **ISO/IEC 8859-5:1999** – *Information technology – 8-bit single-byte coded graphic character sets – Part 5: Latin/Cyrillic alphabet.*
- **ISO/IEC 8859-6:1999** – *Information technology – 8-bit single-byte coded graphic character sets – Part 6: Latin/Arabic alphabet.*
- **ISO 8859-7:1987** – *Information processing – 8-bit single-byte coded graphic character sets – Part 7: Latin/Greek alphabet.*
- **ISO/IEC 8859-8:1999** – *Information technology – 8-bit single-byte coded graphic character sets – Part 8: Latin/Hebrew alphabet.*
- **ISO/IEC 8859-9:1999** – *Information technology – 8-bit single-byte coded graphic character sets – Part 9: Latin alphabet No. 5.*
- **ISO/IEC 8859-10:1998** – *Information technology – 8-bit single-byte coded graphic character sets – Part 10: Latin alphabet No. 6.*
- **ISO/IEC 8859-11:2001** – *Information technology – 8-bit single-byte coded graphic character sets – Part 11: Latin/Thai alphabet.*
- **ISO/IEC 8859-13:1998** – *Information technology – 8-bit single-byte coded graphic character sets – Part 13: Latin alphabet No. 7.*
- **ISO/IEC 8859-14:1998** – *Information technology – 8-bit single-byte coded graphic character sets – Part 14: Latin alphabet No. 8 (Celtic).*
- **ISO/IEC 8859-15:1999** – *Information technology – 8-bit single-byte coded graphic character sets – Part 15: Latin alphabet No. 9.*
- **ISO/IEC 8859-16:2001** – *Information technology – 8-bit single-byte coded graphic character sets – Part 16: Latin alphabet No. 10.*
- **ISO 10585:1996** – *Information and documentation – Armenian alphabet coded character set for bibliographic information interchange.*
- **ISO 10586:1996** – *Information and documentation – Georgian alphabet coded character set for bibliographic information interchange.*
- **ISO/IEC 10646-1:2000** – *Information technology – Universal Multiple-Octet Coded Character Set (UCS) – Part 1: Architecture and Basic Multilingual Plane.*
- **ISO/IEC 10646-1:2000/Amd 1:2002** – *Mathematical symbols and other characters.*
- **ISO/IEC 10646-2:2001** – *Information technology – Universal Multiple-Octet Coded Character Set (UCS) – Part 2: Supplementary Planes.*

Annex D – Some standards from ISO/IEC JTC 1 / SC 32

This list of International Standards relates to clause 2.2.6 of this report.

- **ISO/IEC 6523-1:1998** – *Information technology – Structure for the identification of organizations and organization parts – Part 1: Identification of organization schemes.*
- **ISO/IEC 6523-2:1998** – *Information technology – Structure for the identification of organizations and organization parts – Part 2: Registration of organization identification schemes.*
- **ISO TR 9007:1987** – *Information processing systems – Concepts and terminology for the conceptual schema and the information base.*
- **ISO/IEC TR 9789:1994** – *Information technology – Guidelines for the organization and representation of data elements for data interchange – Coding methods and principles.*
- **ISO/IEC 10032:1995** – *Information technology – Reference model of data management.*
- **ISO/IEC 11179-1:1999** – *Information technology – Specification and standardization of data elements – Part 1: Framework for the specification and standardization of data elements.* – Under revision; new title: *Information technology – Data management and interchange – Metadata registries (MDR) – Part 1: Framework for the specification and standardization of administered items.*
- **ISO/IEC 11179-2:2000** – *Information Technology – Specification and standardization of data elements – Part 2: Classification for data elements.* – Under revision; new title: *Information technology – Data management and interchange – Metadata registries (MDR) – Part 2: Classification for administered items.*
- **ISO/IEC 11179-3:1994** – *Information technology – Specification and standardization of data elements – Part 3: Basic attributes of data elements.* – Under revision; new title: *Information technology – Data management and interchange – Metadata registries (MDR) – Part 3: Registry metamodel (MDR3).*
- **ISO/IEC 11179-4:1995** – *Information technology – Specification and standardization of data elements – Part 4: Rules and guidelines for the formulation of data definitions.* – Under revision; new title: *Information technology – Data management and interchange – Metadata registries (MDR) – Part 4: Rules and guidelines for the formulation of data definitions.*
- **ISO/IEC 11179-5:1995** – *Information technology – Specification and standardization of data elements – Part 5: Naming and identification principles for data elements.* – Under revision; new title: *Information technology – Data management and interchange – Metadata registries (MDR) – Part 5: Naming and identification principles for administered items.*
- **ISO/IEC 11179-6:1997** – *Information technology – Specification and standardization of data elements – Part 6: Registration of data elements.* – Under revision; new title: *Information technology – Data management and interchange – Metadata registries (MDR) – Part 6: Registration of administered items.*
- **ISO/IEC 13249-1:2000** – *Information technology – SQL multimedia and application packages – Part 1: Framework.* – Under revision.
- **ISO/IEC 13249-2:2000** – *Information technology – SQL multimedia and application packages – Part 2: Full-text.* – Under revision.
- **ISO/IEC 13249-3:1999** – *Information technology – SQL multimedia and application packages – Part 3: Spatial.* – Under revision.
- **ISO/IEC 13249-5:2001** – *Information technology – SQL multimedia and application packages – Part 5: Still image.* – Under revision.
- **ISO/IEC FCD 13249-6** – *Information technology – Database languages – SQL multimedia and application packages – Part 6: Data mining.*
- **ISO/IEC 14957:1996** – *Information technology – Representation of data elements values: Notation of the format.*
- **ISO/IEC TR 15452:2000** – *Information technology – Specification of data value domain.* – Under revision.
- **ISO/IEC FCD 15944-1** – *Information technology – Business agreement semantic descriptive techniques – Part 1: Business operational aspects of open-edi for implementation.*
- **ISO/IEC AWI 15944-2** – *Information technology – Business agreement semantic descriptive techniques – Part 2: Registration of scenarios and their components.*
- **ISO/IEC AWI 15944-3** – *Information technology – Business agreement semantic descriptive techniques – Part 3: Open-edi description techniques.*
- **ISO/IEC AWI 15944-4** – *Information technology – Business agreement semantic descriptive techniques – Part 4: Business transaction scenarios – Accounting and economic ontology.*
- **ISO/IEC NP 18022** – *Information technology – Identification, mapping and IT-enablement of existing standards for widely used encodable value domains.*

- **ISO/IEC AWI 18038** – *Information technology – Identification and mapping of various categories of jurisdictional domains.*
- **ISO/IEC FPDTR 20943-1** – *Information technology – Procedure for achieving data registry content consistency – Data elements.*
- **ISO/IEC AWI 20943-2** – *Information technology – Procedure for achieving data registry content consistency – XML structured data.*
- **ISO/IEC AWI 20943-3** – *Information technology – Procedure for achieving data registry content consistency – Value domains.*
- **ISO/IEC AWI 20944-01** through **09** – *Information technology – Metadata registry – Bindings – Parts 1 through 9: Conformance; Language independent datatypes; XML; ASN.1; Language independent procedure calls; C; C++; Java; Javascript.*

Annex E – Some standards from ISO/IEC JTC 1 / SC 35

This list of International Standards relates to clause 2.2.7 of this report.

- **ISO/IEC 9995-1:1994** – *Information technology – Keyboard layouts for text and office systems – Part 1: General principles governing keyboard layouts.*
- **ISO/IEC 9995-2:1994** – *Information technology – Keyboard layouts for text and office systems – Part 2 : Alphanumeric section.* – Amended 1999; under revision.
- **ISO/IEC 9995-3:1994** – *Information technology – Keyboard layouts for text and office systems – Part 3 : Complementary layouts of the alphanumeric zone of the alphanumeric section.* – Amended 1999; under revision.
- **ISO/IEC 9995-4:1994** – *Information technology – Keyboard layouts for text and office systems – Part 4 : Numeric section.* – Under revision.
- **ISO/IEC 9995-5:1994** – *Information technology – Keyboard layouts for text and office systems – Part 5: Editing section.*
- **ISO/IEC 9995-6:1994** – *Information technology – Keyboard layouts for text and office systems – Part 6: Function section.*
- **ISO/IEC 9995-7:1994** – *Information technology – Keyboard layouts for text and office systems – Part 7 : Symbols used to represent functions.* – Amended 1996; under revision.
- **ISO/IEC 9995-8:1994** – *Information technology – Keyboard layouts for text and office systems – Part 8: Allocation of letters to the keys of a numeric keypad.*
- **ISO/IEC 11581-1:2000** – *Information technology – User system interfaces and symbols – Icons symbols and functions – Part 1: Icons – General.*
- **ISO/IEC 11581-2:2000** – *Information technology – User system interfaces and symbols – Icons symbols and functions – Part 2: Object icons.*
- **ISO/IEC 11581-3:2000** – *Information technology – User system interfaces and symbols – Icons symbols and functions – Part 3: Pointer icons.*
- **ISO/IEC CD 11581-4** – *Information technology – User system interfaces and symbols – Icons symbols and functions – Part 4: Control icons.*
- **ISO/IEC FCD 11581-5** – *Information technology – User system interfaces and symbols – Icons symbols and functions – Part 5: Tool icons.*
- **ISO/IEC 11581-6:1999** – *Information technology – User system interfaces and symbols – Icons symbols and functions – Part 6: Action icons.*
- **ISO/IEC FCD 13251** – *Information technology – Collective standard – Graphical symbols for office equipment.*
- **ISO/IEC 14754:1999** – *Information technology – Pen based interfaces – Common gesture for text editing with pen-based systems.*
- **ISO/IEC 14755:1997** – *Information technology – Input methods to enter characters from the repertoire of ISO/IEC 10646 with a keyboard or other input device.*
- **ISO/IEC 15411:2000** – *Information technology – Segmented keyboard layouts.*
- **ISO/IEC 15412:2000** – *Information technology – Portable computer keyboard layouts.*
- **ISO/IEC CD TR 15440** – *Information technology – Technical report on future keyboards and other associated input devices and related entry methods.*
- **ISO/IEC 18021:2002** – *Information technology – User interfaces for mobile tools for management of database communications in a client-server model.*
- **ISO/IEC FCD 18035** – *Information technology – Icon symbols and functions for controlling multimedia applications.*
- **ISO/IEC CD 18036** – *Information technology – Icon symbols and functions for World Wide Web browsers.*
- **NWI** – *Survey of existing icons and symbols for elderly and disabled persons.*
- **NWI** – *Requirements concerning icons and symbols in IT for elderly and disabled persons.*
- **NWI** – Assessment of cultural and linguistic adaptability in software products.

Annex F – Some standards from ISO / TC 46

This list of International Standards relates to clause 2.3.3 of this report.

- **ISO 9:1995** – *Information and documentation – Transliteration of Cyrillic characters into Latin characters – Slavic and non-Slavic languages*
- **ISO 233:1984** – *Documentation – Transliteration of Arabic characters into Latin characters*
- **ISO 233-2:1993** – *Information and documentation – Transliteration of Arabic characters into Latin characters – Part 2: Arabic language – Simplified transliteration*
- **ISO 233-3:1999** – *Information and documentation – Transliteration of Arabic characters into Latin characters – Part 3: Persian language – Simplified transliteration*
- **ISO/DIS 233-4** – *Information and documentation – Transliteration of Arabic characters into Latin characters – Part 4: Phonemic transcription of Persian*
- **ISO 259:1984** – *Documentation – Transliteration of Hebrew characters into Latin characters*
- **ISO 259-2:1994** – *Information and documentation – Transliteration of Hebrew characters into Latin characters – Part 2: Simplified transliteration*
- **ISO 843:1997** – *Information and documentation – Conversion of Greek characters into Latin characters*
- **ISO 999:1996** – *Information and documentation – Guidelines for the content, organization and presentation of indexes.* – This project relates to alphabetical ordering and has some CDICT relevance.
- **ISO 2108:1992** – *Information and documentation – International standard book numbering (ISBN).* – The standard is currently being revised.
- **ISO 2709:1996** – *Information and documentation – Format for Information Exchange.*
- **ISO 3166-1:1997** – *Codes for the representation of names of countries and their subdivisions – Part 1: Country codes.*
- **ISO 3166-2:1998** – *Codes for the representation of names of countries and their subdivisions – Part 2: Country subdivision code.*
- **ISO 3166-3:1999** – *Codes for the representation of names of countries and their subdivisions – Part 3: Code for formerly used names of countries.* – All three parts of this standard are relevant. It is being continuously maintained; see Maintenance Agency home page: http://www.iso.org/iso/en/prods-services/iso3166ma/index.html.
- **ISO 3297:1998** – *Information and documentation – International standard serial number (ISSN).*
- **ISO 3602:1989** – *Documentation – Romanization of Japanese (kana script)*
- **ISO 3901:2001** – *Information and documentation – International Standard Recording Code (ISRC).* – This, and the other "numbering standards" (ISBN, ISSN, etc.) may have some peripheral interest for CDICT.
- **ISO 5426-2:1996** – *Information and documentation – Extension of the Latin alphabet coded character set for bibliographic information interchange – Part 2: Latin characters used in minor European languages and obsolete typography*
- **ISO 5963:1985** – *Documentation – Methods for examining documents, determining their subjects, and selecting indexing terms.* – Relevant for multilingual indexing and search.
- **ISO 5964:1985** – *Documentation – Guidelines for the establishment and development of multilingual thesauri.*
- **ISO 6861:1996** – *Information and documentation – Glagolitic alphabet coded character set for bibliographic information interchange*
- **ISO 7098:1991** – *Information and documentation – Romanization of Chinese*
- **ISO 7154:1983** – *Documentation – Bibliographic filing principles.*
- **ISO 8459** – *Documentation – Bibliographic data element directory.* Parts 1–5 (1988–1998).
- **ISO 8957:1996** – *Information and documentation – Hebrew alphabet coded character sets for bibliographic information interchange*
- **ISO 9984:1996** – *Information and documentation – Transliteration of Georgian characters into Latin characters*
- **ISO 9985:1996** – *Information and documentation – Transliteration of Armenian characters into Latin characters*
- **ISO 10444:1994** – *Information and documentation – International standard technical report number (ISRN).*
- **ISO 10754:1996** – *Information and documentation – Extension of the Cyrillic alphabet coded character set for non-Slavic languages for bibliographic information interchange*
- **ISO 11940:1998** – *Information and documentation – Transliteration of Thai*
- **ISO/DIS 11940-2** – *Information and documentation – Transliteration of Thai characters into Latin characters – Part 2: Simplified transcription of Thai language*

- **ISO/TR 11941:1996** – *Information and documentation – Transliteration of Korean script into Latin characters*
- **ISO 10957:1993** – *Information and documentation – International standard music number (ISMN).*
- **ISO/DIS 15511.2** – *Information and documentation – International Standard Identifier for Libraries and Related Organizations (ISIL).*
- **ISO/FDIS 15706** – *Information and documentation – International Standard Audiovisual Number (ISAN).*
- **ISO 15707:2001** – *Information and documentation – International Standard Musical Work Code (ISWC).*
- **ISO 15919:2001** Information and documentation – Transliteration of Devanagari and related Indic scripts into Latin characters
- **ISO/DIS 15924** – *Information and documentation – Code for the representation of names of scripts.* – This is a central document that is closely related to ISO 639 (see under ISO / TC 37).
- **ISO/CD 21047** – *Information and documentation – International Standard textual work code (ISTC).*
- **ISO/AWI 21127** – *Computer applications in information and documentation – A conceptual reference model for the interchange of cultural heritage information.*
- **ISO/CD TR 21449** – *Content Delivery and Rights Management: Functional requirements for identifiers and descriptors for use in the music, film, video, sound recording and publishing industries.*
- **ISO/AWI 22997** – *Records management relationships with knowledge management.*
- **ISO/AWI 22998** – *Implementation of access rules in records management.*
- **ISO/AWI 22999** – *Records management relationships with document management.*
- **ISO/AWI 23012** – *Records management relationships with other information management practices.*
- **ISO/AWI 23081** – *Metadata for records and records management processes.*
- **ISO 23950:1998** – *Information and documentation – Information retrieval (Z39.50) – Application service definition and protocol specification.*