Project ref. no.	IST-1999-10647
Project title	ISLE Natural Interactivity and Multimodality Working Group

Deliverable status	Public
Contractual date of delivery	30 September 2002
Actual date of delivery	February 2003
Deliverable number	D8.2
Deliverable title	Guidelines for the Creation of NIMM Data Resources
Type	Report
Status & version	Final
Number of pages	50
WP contributing to the deliverable	WP8
WP / Task responsible	NISLab
Editors	
Authors	Malene Wegener Knudsen, Niels Ole Bernsen, Laila Dybkjær, Thomas Hansen, Valerie Mapelli, Jean-Claude Martin, Niklas Paulsson, Catherine Pelachaud, Peter Wittenburg
EC Project Officer	Philippe Gelin
Keywords	Natural interactivity, multimodality, data resources, standards
Abstract (for dissemination)	This ISLE Deliverable 8.1 from the ISLE Natural Interactivity and Multimodality (NIMM) Working Group has two main parts concerning issues involved in NIMM data resource creation: (i) general resource specifications and (ii) data specifications and documentation. The first part describes issues such as legal aspects, rights, modalities used, organisation of the data, and data recordings in terms of procedures, equipment, recruitment, and assessment. The second part describes specification and documentation of data presented in three different generic modalities, i.e. audio, static image, and video data. An introduction to validation and validation criteria is also provided. Finally the report describes some relevant national and international bodies and organisations involved in data resource creation, dissemination, and documentation.

ISLE Natural Interactivity and Multimodality Working Group Deliverable D8.2

Guidelines for the Creation of NIMM Data Resources

February 2003

Authors:

Malene Wegener Knudsen[1], Niels Ole Bernsen[1], Laila Dybkjær[1], Thomas Hansen[1], Jean-Claude Martin[2], Valerie Mapelli[3], Niklas Paulsson[3], Catherine Pelachaud[4], Peter Wittenburg[5]

1: NISLab, University of Southern Denmark. 2: LIMSI-CNRS, Orsay, France. 3: ELRA, Paris, France.

4: DIS, University of Rome, Italy. 5: MPI, Nijmegen, The Netherlands

Contents

1. Introduction 1

2. General specification 4

2.1 Legal aspects 4

2.2 Rights 4

2.3 Subject recruitment and management 5

2.4 Existing standards to consider 6

2.4.1 Audio Encoding Standards 7

2.4.2 Audio Formats 7

2.4.3 Video Encoding Standards 7

2.4.4 Video Formats and APIs 8

2.4.5 Annotation Standards 8

2.5 Data repository structure 8

2.5.1 Directory structure 9

2.5.2 File nomenclature 9

2.5.3 Encoding and schema files 10

2.5.4 Documents to include 10

2.6 Data repository design and collection 11

2.6.1 Recording hardware 11

2.6.2 Recording site 11

2.6.3 Recording conditions 11

2.6.4 Subject recruitment 11

2.6.5 Recording procedure 11

2.6.6 Assessment and quality control 11

2.6.7 Post-processing 11

2.7 Validation 12

3. Data specification and documentation 13

3.1 Audio data 13

3.1.1 Database contents definition 13

3.1.2 Design of prompting and prompt-sheet 13

3.1.3 Data format 13

3.1.4 Annotation 13

3.1.5 Lexicon 14

3.2 Image data 15

3.2.1 Data format 15

3.3 Video data 15

3.3.1 Data format 15

3.4 Multimodal content 15

3.4.1 Background 15

3.4.2 Modalities 15
3.4.3 Body parts 16
3.4.4 Distractors 16
3.4.5 Markers 16
3.4.6 Interactive media 16
3.4.7 Applications 16
4. Relevant bodies and activities............ 18
4.1 COCOSDA 18
4.2 DCMI 18
4.3 ELRA 19
4.4 IMDI Metadata Initiative 19
4.5 ISO TC37 SC4 19
4.6 LDC 20
4.7 OLAC 20
4.8 SPEX 20
5. Conclusion 22
6. References 24
6.1 ISLE NIMM Working Group reports and website 24
6.2 Other literature references 24
6.3 Other website references 25
7. Appendix 1: Sample sheets 26
7.1 Sample instruction sheet 26
7.2 Sample prompt sheet 28
7.3 Recording waiver (adults) 30
7.4 Recording waiver (children) 31
7.5 Description form – Speech 32
7.6 Description form – Multimodal 36
7.7 Contract for distribution 42

1. Introduction

This document, ISLE NIMM (Natural Interactivity and Multimodality) Working Group report D8.2, is a sequel to the ISLE NIMM Working Group Report D8.1 [Knudsen et al. 2002] which surveyed Natural Interactivity and Multimodality Data Resources world-wide, Current and Future User Profiles, and Markets and User Needs. Like all other ISLE NIMM Working Group reports, ISLE Report D8.1 can be downloaded from http://isle.nis.sdu.dk

ISLE NIMM report D8.1 shows that although NIMM data resources already exist in large numbers and variety, there is no standard yet for how to create and document these resources. Based on the comprehensive survey in D8.1, the present report takes early steps towards providing (i) a detailed format for describing multimodal and natural interactivity databases with the aim of contributing to future standards for such descriptions, and (ii) a set of methodological guidelines for the creation and documentation of new multimodal resources.

Within ISLE and due to a close collaboration between WP 8 and WP10 a metadata description schema (IMDI) was developed in particular for the purpose of discovering and managing multimedia/multimodal resources in the Internet and as an entry point for the Semantic Web [Wittenburg 2002]. Necessarily such a description may not be exhaustive for various well-known reasons. This document is intended to give a broader outline of descriptive elements that may be useful to extensively document all facets of NIMM resources. Such a description can be useful for projects with specific goals and needs as a reference. Since natural interactivity and multimodal resources are a fairly new phenomenon as are the description schemas such as IMDI, we expect that the standards will further develop. This was the reason for the foundation of the ISO TC37/SC4 committee about Terminology and Management of Language Resources, cf. Section 4.5. It is assured by personal involvement that the results of the ISLE work will serve to structure and facilitate the discussions.

In the present context, *natural interactivity* means natural interactive communication, i.e. communication, or exchange of information, which uses the full set of means for conveying information used by humans in situated information exchange in which humans communicate about anything whilst sharing time, location, and perceivable physical context. These means (or modalities, see below) include not only speech but also gaze, facial expression, gesture, body posture, use of referenced objects and artefacts during communication, etc. Natural interactive communication is, by nature, *multimodal*. For instance, if speech is considered a single modality in natural interaction, then gaze may be considered a second, different modality, facial expression a third, etc. Moreover, when communicating with machines, humans may be required to use modalities which have no correspondence in natural interactive human-human communication, such as mouse click codes. When communicating with other humans, people use touch only modestly in order to exchange information, and they do not use anything like touch codes for this purpose, except for exchanging information with the severely disabled. For this reason, multimodality constitutes a wider area of modalities for information representation and exchange than does natural interactivity [Bernsen 2002].

As for data resources (or databases), *de facto* standards for documenting *speech* resources are already rather well developed, as witnessed by, for instance, the long series of European projects which have collected and documented large speech databases according to agreed-upon standards, including SAM (http://www.icp.inpg.fr/Relator/standsam.html), SPEECHDAT (http://www.speechdat.org/), ONOMASTICA (http://www.hltcentral.org/projects/detail.php?acronym=ONOMASTICA), SPEECHDAT CAR (http://www.speechdat.org/SP-CAR/), SPEECON (http://www.speecon.com/), and which all, more or less, have built on one another in order to develop comprehensive description formats for speech resources. The result so far is rather sophisticated speech database guidelines sets which, we advise, could profitably be consulted by anyone who wants to undertake to create not only speech databases for new languages, interactive human-system communication tasks, or otherwise, but also natural interactive communication and multimodal resources more generally. As far as we are aware, the latest set of such guidelines can be found under public deliverables at the SPEECON website. As the reader will notice when reading the present report, these guidelines and formats have served as a starting point for venturing into the much less familiar territory of NIMM database guidelines and formats.

The growing need for documentation standards for NIMM data resources is a function of the growing need for those resources themselves. Thus, probably the main reason why there is now a strong need to extend existing, speech-oriented guidelines for documenting communication resources is that, recently, spoken dialogue has come to be viewed by systems developers for what it is, namely, as only a part, albeit a fundamental part, of full natural interactive communication and often of multimodal communication as well. Today's researchers and industrial developers in human(s)-machine and human(s)-machine-human(s) communication are beginning to venture far beyond spoken dialogue systems into more or less full natural interactive systems development. Like the maturing field of speech applications, the wider field of natural interactive systems development needs high-quality, well-documented data resources in order to prosper. Moreover, far from all researchers and developers in this massively expanding field come from speech applications development or are familiar with speech resources creation and documentation. Rather, they come from many different areas of research and development other than spoken language dialogue systems, such as machine vision, computer graphics, telecommunications, and otherwise, and they do not necessarily bring with them a baggage of knowledge about data resources documentation in the speech field or elsewhere. This report is intended to provide some initial orientation for colleagues who are new to data resources for natural interactive and multimodal systems research and development.

Communication data resources development and exploitation is, fundamentally, a relatively high-cost/high-benefit area. High-quality resources are necessary to stay at the forefront of research and development. Such resources are costly develop and hence tend to be costly to purchase. It is in this context that appropriate resource documentation plays a vital role.

Standards and guidelines for communication data resources documentation is becoming a vital factor in ensuring resource usability and re-usability not only by users other than the resource developers but also by the resource developers themselves. Let us briefly mention some of the reasons.

Resource retrieval. The easiest way to get hold of the high-quality communication data resources one needs for some research or development project is to get them from somebody else for free or, if that is not possible, to buy them. Doing so, however, assumes that it is possible to find relevant resources in order to decide that they fit one's purpose. With the semantic web movement, resource documentation is becoming vital for efficient retrieval of relevant resources. During the work on the ISLE NIMM Working Group report D8.1 which surveys a large number of natural interactivity and multimodal resources world-wide, a large amount of work was spent on simply finding the resources in the first place in order to subsequently evaluate them to decide if they met our criteria for inclusion in the report. For the resources which met those criteria, we developed a simple template for describing them. Filling that template for each identified NIMM resource proved to be as much work as identifying the resources in the first place, due to the lack of standard practice even in simple and general resource documentation. The ISLE NIMM D8.1 resource description template served the purpose for which it was created, i.e. to collect general and comparable information about a large number of data resources. However, it is of course too simple for the requirements of serious users of the resources, who need to know much more details about a resource before deciding if it fits their purposes. Still, is may serve to support a first filtering of potentially relevant resources from potentially irrelevant ones. Following the filtering process, the only viable option in most cases is to contact the resource creators and hope that they can provide the additional information needed to decide if the resource fits one's purposes. Had the resources been fully documented according to world-wide standards, hope and uncertainty would have been superseded by straightforward certainty.

As already indicated, the work on the IMDI schema (http://www.mpi.nl/ISLE/) followed a complementary approach in so far as it carried out many discussions with data providers and users to develop a descriptor set and appropriate vocabularies useful for easy discovery and management, but not requiring too much additional work. For special purposes this schema may not be sufficient. In particular the work in D8.2 was one of the motivations to include flexibility in the IMDI set in form of key-value pairs.

Resource creation. If there do not seem to be any NIMM data resources out there which fits one's needs, it may be preferable to develop new resources rather than use inappropriate existing resources. In this case, the naive resource developers first inclination is to simply develop the resource and then use it, ignoring all but the simplest documentation. What normally happens, as we found talking to the developers of the NIMM resources surveyed in ISLE NIMM D8.1, is that the resource developer first spends large amounts of work creating the resource and ends up contemplating the many things which should have be done differently in order to create an optimal resource for the purposes leading to resource creation. We conclude that, had resource documentation guidelines and standards been available, resource developers could have developed their resources in conformance with those standards and guidelines, avoiding worries later on about what should have been done differently. Thus, comprehensive and standardised resource documentation is not only crucial for other resource users, the documentation guidelines are also essential for developers of new resources. By making explicit the properties of a resource which should be documented, the guidelines serve as guidelines for what the developer should take into account before and during resource development.

Difficulty. Is resource retrieval and creation difficult to do? The answer seems to be that resource retrieval and creation is complex and full of pitfalls, which is why the speech resources field has spent decades refining its guidelines for resource creation and evaluation. It is not difficult to do if one knows what one is doing. If one does not know what one is doing, it is very easy and tempting to ignore the complexity involved. Working towards documentation standards is probably the most efficient way to avoid time-waste, law suits, and other hazards.

What this report offers is far from being a complete guide to NIMM data resources documentation. For one thing, the field is far from being mature enough for standardising the documentation. For another, data resources documentation, even in the field of spoken language dialogue data resources, will always remain, to some extent, a matter of individual purpose. This means that it will probably never be possible to provide complete guidelines for the documentation of speech resources or NIMM resources more generally. It will always be necessary to allow for documentation entries which address particular aspects of the resource, its properties or the circumstances of its creation, which no standard could anticipate. Still, we hope that the discussion in the following will help clarify some of the issues involved in documenting NIMM data resources, enabling the resource creator to avoid some of the many pitfalls in resource creation, enabling the documentor to avoid wasting time reflecting on what needs to be described, and encouraging everyone who has developed a NIMM data resource to approach the issue of data resource documentation as something which needs to be done very carefully because not only the developer but also subsequent users of the resource are likely to encounter unnecessary problems and risks when the documentation is missing or inadequate. It should be noted already here that the non-speech NIMM data resources area is quite large and goes beyond the collective hands-on experience available in the current ISLE NIMM Working Group which is the reason why guidelines with respect to those aspects that are specific to non-speech NIMM data resource creation have not really been developed in any detail in this report, cf. also Chapter 5.

This report has two main parts: general resource specifications and data specifications and documentation. The first part (Chapter 2) describes legal aspects, rights, modalities used, organisation of the data, and data recordings in terms of procedures, equipment, recruitment, and assessment. Chapter 2 draws upon ELRA's long experience in this field. The second part (Chapter 3) describes data presented in three different generic modalities, i.e. audio, static image, and video data, following the structure of ISLE NIMM D8.1. Focus is on audio data. An introduction to validation and validation criteria is also provided. Chapter 4 describes some relevant national and international bodies and organisations involved in data resource creation, dissemination, and documentation. Chapter 5 presents brief conclusions. Chapter 6 provides references, and the Appendix (Chapter 7) presents a series of sample resource documentation-related documents all coming from ELRA.

2. General specification

2.1 Legal aspects

The legal aspects primarily concern the subjects recruited for the data recordings. Most often, a written consent to their participation in a recording is required. Also, if minors are included in the recordings, written consent of their parents should be obtained. It is important to note that consent not only concerns subjects' consent to being recorded but also their consent to the intended future use of the recorded data. Resource creators should be aware that, depending on the nature of the data to be collected and its intended use, legal issues could be involved and should be taken very seriously. A simple example is the fact that most parents today are likely to be concerned if stills or videos of their children end up on some website, even if anonymised and even if the purpose is simply to illustrate research. The specificity of, e.g., NIMM video corpora, means that it is more difficult to ensure anonymity of these resources than is the case for speech-only corpora. Moreover, data creators are advised that the legal rights concerning anonymity of computerised data differ from one country to another. Users should be asked their authorization for any kind of use that will be forecast (e.g. internal/research use, distribution/commercialisation within R&D products) Sample recording waivers can be found in Appendix 1.

When recordings are being made outside the company perimeter, authorisation may be needed. This concerns, in particular, public spaces such as train stations, parks, shopping malls, etc. It may therefore be necessary to contact the authorities concerned to obtain authorisation and also to discuss a suitable recording spot, security issues, etc. This is important in order to blend into the environment and to not disturb the usual clients visiting the location.

Authorisations for recordings outside the company are especially important when collecting multimodal data including photo or video capture. In most public places, whether it be the street or a train station, an authorisation for filming is required.

The legal aspects of distributing the database should be considered from early on to ensure that data and agreement forms do not conflict with the plans for later distribution of the created resource. Sample contracts for distribution can be found in the Appendix. An alternative is to leave the legal aspects as well as the distribution to an agency, such as ELDA, the European Language Resources Distribution Agency, cf. Section 4.3.

International laws about privacy are very different, and increasingly often ethical behaviour is required (code of conducts). The European law states that informant signatures are relevant only when they are fully aware of the consequences. Thus: a consent for a distribution via the web is only relevant if the person knows exactly what the web is.

2.2 Rights

In the resource documentation it is in the resource creator's interest to ensure that the intellectual property rights relating to the resource are fully documented. In the interest of furthering scientific progress, it is recommended that it be stated clearly that the resource may be used for research purposes.

In order to negotiate the rights for further distribution, ELDA drafted models of Distribution agreements allowing the distribution of Language Resources to the language engineering community to be adjusted and submitted to NIMM providers. This Distribution agreement is given in the Appendix.

Distribution contract issued by ELDA must be signed by the legal representative of the rights holder organization.

Authorizations should have, where possible, a contractual form and it will be given in an irrevocable (i.e. non-withdrawn) form.

The authorization generally must cover distribution for scientific and commercial goals within the limits of research and development in the field of NIMM technology. It should be explicitly ensured that data will never be used for goals competing with the content-based data.

It should be noted that according to European law every instance that adds value to a resource has copyrights on the resource.

2.3 Subject recruitment and management

There does not appear to be any obvious reason why the recruitment and management of subjects would be significantly different in the case of natural interactivity and multimodal corpora compared to the case of speech resource creation.

Depending on the purpose of the resource, subjects may have to be recruited according to many different specifications and their number may range from a single subject to thousands, in which (latter) case subject planning and management becomes a substantial task in itself. Possible criteria include, among others, age group, gender, weight, height, smoking and drinking habits, information region/dialect, accent, native speakers, non-native speakers, naive or expert users, users with specific backgrounds (knowledge, expertise, education, profession, etc.), professional or untrained subjects, socio-linguistic information, social background, cultural background, communicative behaviour, such as subjects who tend to use gestures a lot, or subjects with particular disabilities, literacy, illiteracy, etc.

Recruitment tends to be rather straightforward if only a few subjects are needed and unless rare qualifications or other subject characteristics are needed. For obvious reasons, mass recruitment is more demanding, the main problem being to recruit sufficient numbers of subjects meeting the selection criteria. Mass-recruitment can be done in several ways: by using an already established list of subjects, if available, by contacting associations, sports clubs, public institutions like the post office or schools, by putting up posters or distributing flyers, by advertising in the media, something which can often be done for free if one has a story with which to attract journalists. In Denmark, for instance, it is possible to purchase lists of people interested in being involved in interviews, polls, etc. from companies that make public polls, such as Gallup. If representativity of the population at large, or of some sub-population, is important, one has to identify people using methods similar to those used by the public polls institutes.

Once recording has started, the subjects may be asked if they know any other people fulfilling the selection criteria. This could evolve into a snowball effect where people will call in at their own initiative as the rumour spreads.

An important point about recruitment has to do with rewards. Usually, it is easier to rally subjects when using a system of rewards for their participation. The reward is supposed to be a compensation for the time spent by the subject on the recording task as well as covering transportation costs. The reward does not necessarily need to be in cash. Cinema cheques, coupons for supermarkets, or similar rewards may be considered as well. In most cases, rewards like these do not seem to be overly dependent on the nature of the subject population sought.

Some frequently used selection criteria are:

- *Age:* this criterion is not too complicated to apply. However, when recording children it can be essential to determine whether or not the speaker has passed the voice break.
- *Region/accent:* the region is sometimes hard to determine since people tend to move around a lot these days. In general, a speaker is considered as belonging to a certain region if he/she has passed the major part of his/her childhood in this region. This definition can be further operationalised in various ways, such as: in which region did you pass most years when aged between 0 and 18 years of age?

Several other possible selection criteria were mentioned above. For all of them, as illustrated in the cases of age and region, a key point tends to be to define and apply some appropriate operationalisation of the criterion, “appropriate” meaning that the operational version of the criterion should be convincing to specialists who would use the data for the purposes at hand.

Other important aspects include:

- the motivation for e.g. selection of particular regions and particular age groups, and the distribution of gender and age per region, and the number of speakers per group should be made clear.

- plan in advance in full detail which information entries are needed on each subject and make sure that the subjects provide the information at the time they are being recorded. This saves time and effort by avoiding that one has to contact the subjects later on in order to obtain the information. An appropriate resource documentation plan should provide the information entries needed;

- maintain a list, or database, of the subjects that have been recorded already in order to avoid that any subject is being recorded twice (in case this is a criterion), be able to contact particular subjects later on in case something went wrong during the recording, etc. As soon as the number of subjects recorded exceeds a few subjects, it is recommended to use a database for managing the information, generating statistics, checking and warning if fixed thresholds of numbers of subjects with specific properties have been recorded, etc. It may even be necessary to ask subjects for an ID to make sure that recordings will not have to be rejected later on if the recorded subjects do not meet the criteria for their participation in the recordings.

- double-check on age, region, and other selection criteria before recording starts;

- make a single person, the subjects supervisor, in charge of recruitment and subject management.

Obviously, these recommendations are but a handful of the tricks of the trade of subjects recruitment and management. More can be found in the bodies of documentation references in this report. However, even when the supervisor is equipped with state-of-the-art documentation templates for guidance, subject management remains an area in which vigilance and pro-activeness is at a premium because no set of guidelines says it all, as the following example shows. Even if the SPEECON project probably is among the best documented speech resource projects in terms of planning and documentation needs, the project guidelines do not include the following gem: in Northern Europe, it is winter time between November and March. So, if your recording plan assumes that a minimum of 50 recordings should be made outdoors in public spaces, avoid, if possible, to make these recordings in winter because you may have to do them for weeks in sub-zero temperatures in some windy public area. Incidentally, this story highlights that subject management may also have to deal with planning the availability of particular recording environments and set-ups (see also below). In another example: if subjects are to be recorded when driving a car, make sure that they have a driver’s license and that appropriate insurance is in place. None of this is likely to be stated even in comprehensive recording guidelines, very likely because these issues do not form mandatory parts of the subsequent documentation.

2.4 Existing standards to consider

For the creation of annotated multimodal resources a number of standards are relevant to be considered. These standards refer not only to the encoding of audio, video and textual material, but also to the encapsulating format of the resulting resources. In a domain of continuously increasing complexity it is necessary to determine the purpose of the resources to be created and the intended longivity. In particular these criteria will determine whether one has to follow established standards or whether one wants to invest in standards that are just emerging. This chapter cannot give a complete overview of the relevant standards but will mention some of the most popular to indicate possible choices to the reader.

2.4.1 Audio Encoding Standards

For speech data there are several different standards depending on the type of application and media. For desktop speech, for instance, a PCM 16kHz or 20 kHz, 16bit linear format is used most often, while an A-law 8kHz, 8 bit format is common for telephony speech. Current DAT and flash-memory recorders record sound at 44.1 kHz stereo with 16 bit linear PCM delivering high quality. However, not all applications require such a high level of quality. Currently, MPEG2 layer3 (short MP3) and MiniDisc (ATRAC)[1] compressed audio formats are very popular, since the devices and media are inexpensive and small. Both compression techniques apply a psycho-acoustic type of filtering [Campbell 2002], i.e. they remove sound characteristics in the frequency and in the time domain that are said to be irrelevant for our acoustic perception. Tests have been carried out by several people [Campbell 2002, Wittenburg 2002] that indicate some effects of these algorithms on the representation of the speech signal. It is widely agreed that for pure short-term recording purposes compressed speech may be an acceptable option. However, for long-term archiving and re-use of speech and even more important for sound in general, the recommendation is to not use compressed speech, but to use the best quality recording one can achieve. This view is supported by the associations of audio archives.

2.4.2 Audio Formats

The most popular file format is WAVE format, since it was adopted by Microsoft. Most audio processing programs support this de facto standard. Another important format was the NIST audio format, which was very well-known in the area of speech engineering labs. There are enough programs that reliably convert between these standards. One has to be aware that the header information that is for example included in the NIST files may partly be lost.

2.4.3 Video Encoding Standards

Here we are confronted with a large variety of compression techniques to reduce the amount of storage needed. The most popular format for semi-professional recording devices is the proprietary DV (Digital Video) format from Sony. It has replaced older analogue recording techniques such as Hi-8. DV creates video streams with a byte rate of about 33 MB/sec. New types of MPEG2 recorders that are compressing moving images according to the MPEG2 algorithms are fairly new on the market and have to show their success.

For computer applications the whole family of MPEG compression techniques are the most relevant ones and completely replaced older formats such as Cinepak (known in the MacIntosh World) and MJPEG (that only did compression on one single frame). MPEG1 still is the most frequently used standard. It can have a variable bit rate. In general bit rates of about 1.0 or 1.5 Mbps are chosen. Given these typical bit rates MPEG1 is still used in applications where video has to be transmitted from CDROM or via a local area network. It could also be used for high bandwidth transmissions (such as offered by XSDL techniques), but it will be replaced here by MPEG4. MPEG1 is not used anymore as archiving or editing format. Its effective resolution is comparable with VHS, while the newer standard MPEG2 offers resolution comparable with S-VHS. Especially for multimodal resources where the video signal will be used to analyse for example facial expressions or the movements of fingers, MPEG2 will be the choice[2]. Due to its easier handling MPEG2 is now also the choice for almost all relevant editing systems. For archiving purposes the large institutions mostly

[1] It should be noted that the ATRAC algorithm is a proprietary format from Sony, while the MP3 algorithm is publicly documented in all details.

[2] Of course, the camera position in relation to the recorded object is of greatest relevance. Often, however, scientists want to capture all movement aspects of the subject with one camera, i.e. a certain distance has to be chosen. In these cases MPEG2 offers substantially more detail than MPEG1.

choose MPEG2 as well. DV is not a choice, since it is proprietary and since its bit-rate is much higher compared to MPEG2 while not offering essentially better resolution.

The newly developed MPEG4 standard is meant to support the decoding and merging of several video objects (streams). It comes along with new compression algorithms for video at different bit-rates. Comparing a 1 Mbps stream MPEG4 offers more information compared to MPEG1. Therefore, the MPEG4 compression algorithms will replace MPEG1 for many applications such as for example video streaming across networks. There are some other compression algorithms such as Sorensen, however, they are not widely used. RealVideo presented its own proprietary video compression standard, since there was nothing available that could stream video across the Internet. It is assumed that these compression techniques will be replaced by MPEG4. There are conversion programs at the market, but due to the complexity of the problem, it seems that the conversion is still error prone. So one has to check which of the programs available can actually do the specific conversion intended in a correct way.

Summarising, we can say that MPEG2 is the preferred encoding standard for archiving and editing. MPEG4 seems to be the best option for video streaming applications in the near future. This is the reason why professional institutions often store a copy in MPEG2 and create (sometimes several) MPEG4 versions dependent on the available network bandwidth.

2.4.4 Video Formats and APIs

There are a few widespread formats for video files such as MPG, AVI and QuickTime. While MPG and AVI are simple container formats, QuickTime can be said to offer even more functionality. While for example the MPEG standards define how video and audio signals are combined into one bitstream, QuickTime also allows to associate many tracks and to interlink them in time. Such additional tracks could for example be textual annotations. We will not discuss these issues in this document in more detail. Again one has to look carefully at the beginning of the project what the purpose is. Depending on the choices one has to select a suitable format. There are programs that can convert the audio/video information between these formats.

2.4.5 Annotation Standards

This note cannot give an overview about the many standards or best practise guidelines of how to encode linguistic phenomena. There is a clear trend to use UNICODE that for example includes the phonetic alphabet as character encoding standard. Increasingly more software is around that can convert from a specific character encoding template to UNICODE. In the same way there is general agreement to use XML as the common language for structuring textual annotations. Right now there are a few schemas (ATLAS (http://www.nist.gov/speech/atlas/), ACM [Brugman 2001]) that allow to cater for flexible annotations covering all structural phenomena that can occur in multimodal resources and that also have mechanisms to link annotations with the media signals. The different proposals are also subject of thorough investigation in the ISO TC37/SC4 committee.

At the level of the encoding of linguistic phenomena we can mention SAMPA as a standard for encoding phonetic/phonemic information. For morphosyntax the EUROTYP encoding scheme has been suggested, however, it was already identified that it was made by having the western type of languages in mind. For annotating prosodics, the TOBI conventions are often applied. See ISLE report D9.2 for more information on standards for annotation schemes.

2.5 Data repository structure

Resources have to be stored such that management and direct access is facilitated. This is especially true for repositories holding a large number of resources of different types and origins. In the Internet era an additional dimension is added since resources bundled by some criteria, such as resulting from the same project, will most likely be stored in a distributed fashion. Another dimension is given by the increasing need to move and copy the content of such repositories for data security or storage media management reasons.

We can distinguish between traditional methods of repository and modern repository structures. Traditional repositories divide their holding into fixed resource bundles which are called “the XYZ corpus”. Access is only given to the whole corpus and also management is done for the corpus as a whole. Modern repositories especially in the Internet era are archived by defining several layers. The physical storage (disk, server) is only known by the system management specialists. According to the needs at this level data is copied and distributed on different servers, storage containers and locations. However, all references to the resources have to be maintained. Such a system requires the introduction of a unique identifier as was observed for example by the DOI initiative. At the user and data manager level all system operations should be completely transparent. To achieve this in modern repositories a metadata layer is introduced that describes the resource by domain specific descriptors. Each resource is associated with such a metadata description and both are linked via the unique resource identifier and a mapping database storing the physical locations of the copies.

In this report we cannot discuss these issues in detail. It is again the ISO TC37/SC4 committee that will look at various proposals made by e.g. ISLE NIMM IMDI to derive methods that match with all future needs. In this note we simple want to raise a few points that indicate the aspects to be thought of at the physical level of the repository organization. It describes one possible way of structuring, however, it is obvious that each project requires its own specifications.

2.5.1 Directory structure

The directory structure being discussed as an example uses a shallow directory nesting with contiguous numbers to identify the individual sub-directories and call directories. The following three-levels directory structure serves as an example and has been taken from speech databases:, see also Figure 2.1.

\<database>\<block>\<session>

The term database refers to a bundle of resources that may have been created in a specific project as part of a whole corpus with specific goals in mind as for example a multimedia corpus with speech and gesture annotations to analyse the timing correlation between these two modalities and to feed statistical engines. While the corpus could be gathered in several countries, the database refers to the part that was gathered in one specific country. The term block identifies some meaningful grouping of resource in the database such as informants sharing the same age. The term session refers to a leaf in the corpus which is the bundle of resources belonging to one specific interview for example. This bundle can include the media and the annotation files for example. For further details we refer to the ISLE IMDI documents (http://www.mpi.nl/ISLE/).

<database>	Defined as <dbName><#><language code>
<block>	Defined as BLOCK<NN>
<session>	Defined as SES<NN><M>

Table 2.1. Directory structure.

2.5.2 File nomenclature

For file naming one has to decide about a consistent scheme such as suggested by the ISO 9660 (http://www.iso.org/iso/en/ISOOnline.frontpage) standard[3] According to ISO 9660 file names have 8 characters followed by a 3-character file extension:

[3] Today in most cases this limited length of filenames cannot be applied anymore.

<dbID><NNM><CCC>.<LL><F>

where:

<dbID>	Database Identification Code (00-ZZ)
<NNM>	Progressive recording session number (000 to 999), where NN is the block number and M is the session number
<CCC>	Corpus code
<LL>	ISO 639 language code
<F>	File type code: O: orthographic label file

Table 2.2. File name conventions.

2.5.3 Encoding and schema files

The corpus should include information about the schemas used to structure the resources and the encodings applied:

- Indication of labels used within the schema and short description of their meaning.
- Information on character encoding used: UTF-8, IPA Times, ...
- Information on encodings used for media: PCM/48/16, MP3, MPEG4, ...
- Information on encoding schemes which were used to describe linguistic phenomena: SAMPA, TOBI, ...
- Example files.

2.5.4 Documents to include

Documents included (ISO codes, tables with statistics, etc.) should be listed.

Documents to be provided are:

- COPYRIGH.TXT: a copyright text in ASCII format,
- DISK.ID: an 11-character string with the volume name (required for systems that cannot read the physical volume label),
- README.TXT: an ASCII text file that lists all files of the database, except for signal and label files which can be indicated by their name template and contains contact information.

These should preferably be placed in the root directory. Other document may be

- tables of statistics,
- documentation,
- description of standards, e.g. list of SAMPA symbols for the given language, etc.

2.6 Data repository design and collection

2.6.1 Recording hardware

This section of the documentation should include the specification of recording platform (e.g. microphone, video or camera equipment, amplifiers, laptops, extra computers for post-processing, VCRs, multi-channel recordings, etc) and reference to an appendix with detailed information about the equipment used.

2.6.2 Recording site

- Description of the site of recording and motivation why it was chosen.

2.6.3 Recording conditions

- Description of different environments (home, office, outside) with detailed explanations of each site.
- Positions during recordings: speaker and microphone positions as well as the movements or positions of impostors or other people.
- If cars or other automotive vehicles are used: descriptions of type of vehicle, conditions (speed, dark, rain, etc), status of windows, etc.
- Short description of background, illumination and scenes.
- Scene: Illumination - daylight, single source, multiple sources, fix, variable. Background – plain, complex, etc.
- Short description of background noise, if relevant.

2.6.4 Subject recruitment

See Section 2.3 for a description of which elements may be relevant.

2.6.5 Recording procedure

- Procedures for the recording: types and positions for all environment types, setup of equipment, etc.
- Recording procedure for one session
- Backup of data, etc.

2.6.6 Assessment and quality control

- Quality assurance and control before, during and after recordings, etc.
- E.g. checks on signal files, statistics, progress tracking, etc.

2.6.7 Post-processing

- Any post processing of data files, annotations, corrections, methods of signal processing, filtering, etc.

2.7 Validation

It is advisable to have the resource validated by an independent validation centre. Currently there is a centre in the Netherlands (SPEX) which ELRA uses for speech resources validation. There is also one in Germany (BAS) for multimodal resources. Competent centres have to be found for validating multimodal resources.

The validation adds to the quality by certifying that the product follows the agreed upon specifications and that any future producers of similar resources follow the same set of specifications. Deviations from specs, e.g. additional optional recording materials, should be mentioned. Usual quality checks for file structure, signal/image quality and enough documentation is performed during this process.

ELRA has created a Validation Committee in order to encourage data producers in the field (see Validation section of the ELRA web site at http://www.elra.info).

3. Data specification and documentation

3.1 Audio data

3.1.1 Database contents definition

3.1.1.1 Linguistic content

- Presentation of corpus: corpus codes, lists of items, etc.
- Description of corpus, e.g. free spontaneous speech, isolated digit strings, continuous digit strings, read speech, times, dates, spelled words, names, phone numbers, natural numbers, phonetically rich words and sentences ensuring that each sound in a language is represented a minimum number of times, application words, isolated words, command vocabulary, yes/no questions/answers, size of corpus, etc.
- Phonetically balanced material including statistics and motivation for selection.

3.1.1.2 Speaker information

- Number of speakers, male/female, imposters, synthetic (avatars), children.
- Distribution of age. Typical distribution for speech applications: (8-11, 12-15, 16-30, 31-45, 46-80). Minimum number of speakers for each age group.
- Origin: native, non-native.
- Geographic distribution: choice of regions, number of speakers per region, dialects.
- Additional information: place of living, dialect, place of birth, secondary education, speaking/hearing impairments, smoking habits, heights, profession, weight, trained subject, etc.

3.1.2 Design of prompting and prompt-sheet

- How many.
- Why.
- Oversampling.
- Motivation for spreading of items over prompt sheet.
- Reference to example of prompt sheet, to be added as an appendix.

3.1.3 Data format

- File formats, encoding, sampling frequency, quantisation, compression, etc.

3.1.4 Annotation

3.1.4.1 Contents

- Standards and references used for producing this resource.
- Spelling standards, description of any deviation to such a standard.

- List of non standard and alternative spellings.
- Character set used for annotation.
- Annotation conventions for digits, numbers, spelt letters, punctuation, use of capital letters.
- Language-dependent information such as abbreviations, proper name conventions, contractions.
- List of symbols for non-speech acoustic events as well as mispronunciations, truncated signals, non-intelligible parts and background noise.
- List of symbols for annotation of modalities (movement, actions, etc).

3.1.4.2 Procedures

- Selection of annotators. Annotators should have some basic training in linguistics and be adept at spelling. Limit the number of annotators to a minimum in order to preserve consistency throughout the entire transcription/annotation process. The more annotators, the greater the likelihood of inconsistencies.
- Training of annotators. Training in transcription and annotation should be done and supervised by a trained linguist who is also a native speaker of the language. Devise small practice segments which can be completed by the annotators.
- Annotation. Annotation should be done in multiple passes. One pass for each modality concerned, in order to limit the risk of mistakes posed by multitasking.
- Description of quality assurance procedures.
- Annotation manuals containing guidelines and instructions. Prior to beginning the annotation process, manuals and guidelines should be worked out in order to preserve consistency. The manuals should contain all imaginable problems and scenarios involved in the process.
- Procedures for double checking annotations. The original annotation supervisor should do spot-checks of the annotations. As new problems or questions arise in the course of the annotation procedure, these should be brought to the attention of the supervisor, solved and the manuals updated accordingly. Furthermore, all annotations should be checked by a second supervisor, who is a trained linguist and a native speaker of the language.

3.1.4.3 Tools

- Software tools used for annotation.

3.1.5 Lexicon

3.1.5.1 Contents

- Format of lexicon.
- Explanation or reference to the phoneme set used.
- Statement whether or not the annotations and the lexicon are case sensitive.
- Standards (SAMPA, DARPA) and list of standard phone symbols.
- List of PinYins and Hepburn Romaji syllables (if applicable).
- Information captured in the phone transcriptions (assimilation and reduction rules).
- Statement whether multiple transcriptions are supported.

- Statement whether stress information is supplied.
- Statement whether there are any tags, and if so, the tagging conventions used, e.g., record (noun) vs. record (verb).
- List of words that are from a foreign language.
- List of rare phonemes.
- Any other language-dependent information or conventions.

3.1.5.2 Procedures

- Procedures to obtain phonemic forms from orthographic input.

3.1.5.3 Statistical information

- Analysis of frequency of occurrence of the phonemes represented in the COMBINED phonetically rich sentences and phonically rich words and in the full database (at transcription level).
- Word frequency tables.

3.2 Image data

3.2.1 Data format

- File formats, image resolution, encoding, sampling frequency, quantisation, compression, file size, etc.

3.3 Video data

3.3.1 Data format

- File formats, image resolution, frame rate, encoding, sampling frequency, quantisation, compression, file size, etc.

3.4 Multimodal content

3.4.1 Background

- Short description of why creating a multimodal resource, objectives and goals.

3.4.2 Modalities

Modalities describe the extra layer of information which is extracted from the resource. This could be body movements, gestures, expressions or manipulation of objects. Most often these modalities are annotated with a coding scheme.

- Reason for choosing modalities, which modalities, statistics and distribution.
- Facial expressions – anger/irritation, boredom, joy, surprise, neutral, etc.
- Head movements – rotation, inclination, etc.
- Face view – frontal, profile, etc.

- Gestures – interactional, non interactional, emotional, etc.
- Gaze/eye movements – saccades, pursuit, convergence, horizontal, vertical, etc.
- Hand manipulation – direct, indirect, modifying objects, joining/splitting, changing position, etc.
- Body movements – upper body, lower body, etc.

3.4.3 Body parts

Body parts describe both the parts of the subjects' body that have been deliberately included and the ones that are not.

- Information about which body parts that are included in the resource and why they have been selected.
- E.g. whole body, head, face, mouth, arms, hands, legs, feet, etc.
- Schemes for movements or procedures for including the body parts in the resource, e.g. lifting the right hand at a certain time or to the response of a certain event.

3.4.4 Distractors

Distractors refer to any object that can hide the whole or parts of the subject, usually the face of the subject. The aim can be either to test a system or to include training data which is not perfect.

- Description of distractors that have been deliberately included and why.
- E.g. hat, glasses, markers, watch, scarf, pen, paper, notepad, microphone, mobile phone, etc.

3.4.5 Markers

Markers are small tags which are used for tracking movements, most often head movements. The markers are placed on, e.g., the face of the subject and can be reflective, responsive to ultraviolet light or colour.

- Use of markers and reference to any annotation schemes for these.

3.4.6 Interactive media

Interactive media is any object or device which can be used for an interactive purpose, such as a computer screen or a data glove in order to ease the communication.

- Description of all interactive media used during the recording and its purpose.
- E.g. graphical screen, computer pen, tactile screen, data glove, PDA, desktop, laptop, mouse, bluetooth, etc

3.4.7 Applications

- Target application areas for this resource: education/training, research, entertainment, banking, tourism, etc.
- Applications.
- Authentication – face verification, speech verification, user authentication, etc.

- Recognition: face recognition, automatic speech recognition, person recognition, expression recognition, etc.
- Analysis: lip tracking, speech/lips correlation, etc.
- Synthesis: talking heads, avatars, humanoids, multimedia, etc.
- Control: voice control, speech assisted video, etc.

4. Relevant bodies and activities

In this section we briefly describe initiatives in the area of natural interactivity and multimodal communication which have set standards for NIMM data resources or are currently seeking to do so. ISLE report D9.2 describes other initiatives which are primarily of interest to NIMM annotation schemes but some of which also touch upon data resources and thus are of relevance to this report. The described initiatives in D9.2 include TEI, ToBI, SAMPA, ISO TC37/SC4, MPEG-4, MPEG-7, ATLAS and NITE [Dybkjær et al. 2003]. In the same way a couple of the projects described in ISLE report D11.1 also mention data resources (ATLAS and TalkBank) [Dybkjær et al. 2001].

4.1 COCOSDA

COCOSDA (The International Committee for the Co-ordination and Standardisation of Speech Databases and Assessment Techniques) (http://www.cocosda.org/) started in 1992. The goal of COCOSDA is to set up, encourage and support international interaction and cooperation in developing spoken language resources and speech technology assessment methodologies. COCOSDA is structured around *Topic Domains* and *Regions.* Currently, COCOSDA is addressing the following four topic domains:

- Evaluation of Speech Understanding/Dialogue Systems,
- Multi-Modal Corpora,
- Corpus Annotation Tools, and
- Local Languages.

New topic domains may be added and old topic domains may be removed or revised whenever relevant. There are six geographical regions: North America, Europe, Asia, Oceania, Latin America and Africa.

COCOSDA promotes its work through annual workshops (at ICSLP and Eurospeech) and through its website.

4.2 DCMI

The DCMI (Dublin Core Metadata Initiative) (http://dublincore.org/) started in 1995. It is an open forum for the development of standards for interoperable online meta-data in support of a broad range of purposes and business models. Anybody who wants to join the DCMI can do so. DCMI activities include:

- Standards development and maintenance, such as organising international workshops and working group meetings directed towards developing and maintaining DCMI recommendations.
- Tools, services, and infrastructure, including the DCMI meta-data registry to support the management and maintenance of DCMI meta-data in multiple languages.
- Educational outreach and community liaison, including developing and distributing educational and training resources, consulting, and coordinating activities within and between other meta-data communities.

The basic idea is that a simple and widely-understood set of elements will promote interoperability amongst heterogeneous meta-data systems and improve resource discovery on the internet. The DCMI defines 15 elements, some qualifiers for refining the elements, and some constraints.

4.3 ELRA

ELRA (European Language Resources Association) (http://www.icp.inpg.fr/ELRA/home.html) was established in Luxembourg as a non-profit organisation in February 1995. ELRA seeks to promote language resources for the Human Language Technology (HLT) sector, and to evaluate language engineering technologies. ELRA pursues its goals by offering a number of services, including:

- Identification of language resources.
- Promotion of the production of language resources.
- Production of language resources (in special cases).
- Validation of language resources.
- Evaluation of systems, products, tools, etc., related to language resources.
- Distribution of language resources.
- Standardisation.

ELRA also conducts market studies in the HLT area from time to time and ELRA is behind the bi-annual LREC conference.

ELRA uses SPEX (see below) as their primary validation centre for spoken language corpora.

4.4 IMDI Metadata Initiative

The IMDI (Isle Meta Data Initiative) metadata initiative (http://www.mpi.nl/ISLE) is part of the ISLE project's NIMM working group activities. It has developed a complete metadata infrastructure for multimedia/multimodal language resources including a metadata schema, vocabularies, registry mechanisms and tools. The results are a product of many discussions with field experts and in particular of two workshops. The IMDI domain is a browsable and searchable domain of distributed metadata records that are used for easy resource discovery and management.

Currently, the version IMDI 3.x is in preparation. It is the consequence of almost two years of experience and new requirements resulting from the ECHO (European Cultural Heritage Online) and INTERA (Integrated European Language Resource Area) projects. To enable interoperability a bridge to DC/OLAC was created that allows metadata harvesting and therefore searching in both domains.

4.5 ISO TC37 SC4

ISO/TC37/SC4 (ISO Technical Committee 37 / Sub-Committee 4) started in May 2002 and does not have a web site yet. The work programme can be found at http://www.korterm.or.kr/gita_link_doc/brochure_new.doc. The ISO/Technical Committee 37 aims to prepare standards by specifying principles and methods for creating, coding, processing and managing language resources, such as written corpora, lexical corpora, speech corpora, dictionary compiling and classification schemes. To achieve its goals, ISO/TC37/SC4 has defined a number of sub-areas for its work:

- Descriptors and mechanisms for language resources.
- Structural content of language.
- Semantic content of multimodal data.
- Discourse-level content.
- Multilingual text representation.

- Lexical databases.
- Workflow of language resource management.

TC37 SC4 will promote international standards through technical reports that cover language resource management principles and methods, as well as various aspects of computer-assisted lexicography and language engineering, including their implementation in a broad array of applications.

4.6 LDC

LDC (Linguistic Data Consortium) (http://www.ldc.upenn.edu/) was founded in 1992 with a grant from the Advanced Research Projects Agency (ARPA). It is now partly supported by a grant from the Information and Intelligent Systems division of the National Science Foundation. LDC is an open consortium of universities, companies and government research laboratories. It creates, collects and distributes speech and text databases, lexicons, and other resources for research and development purposes.

Core activities of LDC include:

- Maintaining data archives, producing and distributing CD-ROMs, arranging networked data distribution, as well as negotiating intellectual property agreements with potential information providers and with would-be members, maintaining relations with other groups around the world who gather and/or distribute linguistic data, hosting occasional workshops, and so on.
- Pre-publication processing of data donated by other groups, the production of small or inexpensive databases, and pilot work on larger projects in advance of other funding.
- Much of the planning and overseeing of specific databases is funded by outside sources.
- Various forms of cost-sharing to make the production of databases funded by outside entities (whether governmental or commercial) more efficient.

Though rooted in the USA, LDC has plenty of international contacts and collaborations.

4.7 OLAC

OLAC (Open Language Archives Community) (http://www.language-archives.org/) was the outcome of a NSF-sponsored workshop on Web-Based Language Documentation and Description which took place in December 2000. OLAC is an application of the Open Archives Initiative (OAI) (http://www.openarchives.org/) to digital archives of language resources. The OAI which was launched in 1999 develops and promotes interoperability standards for digital archives, and currently spans dozens of archives and a total of over a million records. OLAC is an international partnership of institutions and individuals who are creating a worldwide virtual library of language resources by:

- Developing consensus on best current practice for the digital archiving of language resources.
- Developing a network of interoperating repositories and services for housing and accessing such resources.

OLAC promotes standards and recommendations for best practice via documents and reports. On the OLAC website there are draft standard and recommendation documents on meta-data and meta-data extensions. The OLAC meta-data set is based on the Dublin Core initiative (see above). The meta-data documents include a definition of the standards that OLAC data providers must follow in implementing a meta-data repository.

4.8 SPEX

SPEX (Speech Processing EXpertise centre) (http://www.spex.nl/) was founded in 1987. Its objective is to develop and provide software, tools and databases for companies and institutes active in research and development in the general field of speech, with an emphasis on speech technology. Being governed by the Dutch Foundation for Speech Technology (SST), SPEX has a special task in making available spoken language resources for research purposes in the Dutch academic environment.

SPEX's main activities are the creation, annotation and validation of spoken language resources, including the following:

- SPEX has been selected as the primary Validation Centre for speech corpora in ELRA.
- SPEX acts as validation centre for several European projects in the SpeechDat framework.
- SPEX is also involved in the annotation and post-processing of human-machine dialogues aimed at developing spoken dialogue systems for information services.

SPEX is thus involved in validation work in several projects, including Speecon and Orientel.

5. Conclusion

Like software, new data resources are often costly and time-consuming to create and document. There are cases, however, in which, e.g., a simple and short video recording of one or a few subjects who could be anyone, such as students of colleagues in the work place, might do for the data analysis purpose at hand. In such cases, nothing should prevent the data resource user from creating the resource since this will be neither costly nor time-consuming to do. It seems likely, however, that many, if not most, of the quality NIMM resources that will be needed in the future will impose stronger requirements on cost and time than that, for instance because they require many subjects, or subjects belonging to different, well-defined groups of which there is no immediate supply, or because the recordings are complex to do or require sophisticated equipment, or because recording scripting is time-consuming to do, etc. Let us call these resources *complex* data resources.

So the first question to ask when there is a need for a quality complex NIMM data resource is: are there any potentially relevant data resources out there already? The problem then becomes one of finding out whether or not this is the case. As shown above, the development of web-searchable meta-data for NIMM resources is in its infancy and therefore cannot presently be relied upon to yield adequate returns. At present, the only alternatives are to consult existing NIMM data resources surveys, such as ISLE NIMM Report D8.1, contact data resource providers, such as ELRA or LDC, ask colleagues using mailing lists or other means, or perform free-form web search.

If, using the means of searching just described, potentially relevant data resources are identified and accessed, the question is whether or not those resources come sufficiently close to meeting one’s needs. Given the generally rather poor state of ready-made data resources documentation found in ISLE D8.1, answering this question is likely to involve analysis of the data resource as well as contacts with its creators. If this process turns up a data resource which is adequate for one’s purposes and which can be accessed and used for those purposes, the search has been successful and one does not have to create any new resource. However, the resource user will often be in a slightly different situation. This situation is one in which the data resources identified are *not quite* to the point. For instance, the non-verbal natural interactive communication behaviours visible in the recording are of sufficient quality and the amount of different behaviours is adequate, but the recording script is a very different one from the one which had been planned. Given the fact that very many data resources are being created for quite specific purposes, such *near-miss situations* would seem likely to arise rather often. In these cases, we believe, many researchers have been inclined to react by modifying their research purpose rather than sticking to it and create their own resource. This is obviously not good for the general progress of research. The message here is simply that we need to make it as easy as possible to create new resources through guidelines and standardisation.

If all attempts to find accessible and focally relevant NIMM data resources fail, the remaining alternative is to create a new resource which is optimised for the purpose(s) at hand. If the resource creator is new to data resource creation, we strongly recommend consulting *existing best practice in speech data resource creation* no matter what is the specific resource creation purpose. This practice constitutes the closest approximation we currently have for guidelines on how to create data resources in the NIMM area without ignoring a large class of important details. In its present state, this report provides large amounts of documentation on best data resource creation practice in the spoken language part of the NIMM data resources area.

It is obvious that spoken language data resource creation guidelines and best practice are not sufficient for the general area of NIMM data resource creation. Although a wealth of experience can be transferred from speech data creation and documentation to NIMM data creation and documentation more generally, all issues to do with non-speech natural interactive and multimodal behaviour scripting, subject instruction, image and video recording, use of specialised equipment for, e.g., facial expression recording, data formats, video signal processing, data post-processing, etc., go clearly beyond the speech field.

Guidelines with respect to those aspects which are specific to non-speech NIMM data resource creation have not been developed in this report. At this point, we must refer the reader to the comprehensive current practice documentation provided in ISLE NIMM Report D8.1. We hope that this situation is temporary and are working towards making it possible for ISLE NIMM Working Group participants and other colleagues to add contributions on this part of the planned D8.2 work in connection with publication of the ISLE NIMM results after the end of the project. The challenge is that the non-speech NIMM data resources area is quite large and goes beyond that collective hands-on experience available in the current ISLE NIMM Working Group.

6. References

6.1 ISLE NIMM Working Group reports and website

The ISLE NIMM (Natural Interactivity and Multimodality) Working Group is one of three working groups in the joint EU/US ISLE (International Standards for Language Engineering) project January 2000 - December 2002. The list below shows the report produced by the ISLE NIMM Working Group. All reports can be downloaded from the ISLE NIMM website (isle.nis.sdu.dk).

Broeder, D., Offenga, F., Willems, D., Wittenburg, P.: Metadata Set for Multimedia/Multimodal Language Resources. ISLE Deliverable D10.2, August 2002a.

Broeder, D., Offenga, F., Willems, D., Wittenburg, P., Heid, U., Vögele, A. and Popescu-Belis, A.: IMDI Showcase. ISLE Deliverable D10.3, September 2002b.

Broeder, D., Offenga, F. and Wittenburg, P.: Overview of Metadata Initiatives and Corpus Metadata in Language Engineering and Linguistics. ISLE Deliverable D10.1, October 2000.

Dybkjær, L., Berman, S., Bernsen, N.O., Carletta, J., Heid, U. and Llisterri, J.: Requirements Specification for a Tool in Support of Annotation of Natural Interaction and Multimodal Data. ISLE Deliverable D11.2, July 2001a.

Dybkjær, L., Berman, S., Kipp, M., Olsen, M.W., Pirrelli, V., Reithinger, N. and Soria, C.: Survey of Existing Tools, Standards and User Needs for Annotation of Natural Interaction and Multimodal Data. ISLE Deliverable D11.1, January 2001b.

Dybkjær, L., Bernsen, N.O., Broeder, D. and Wittenburg, P.: Introduction to and Summary of the Final NIMM WG Guidelines. ISLE Deliverable D7.1, February 2003.

Dybkjær, L., Bernsen, N.O., Knudsen, M.W., Llisterri, J., Machuca, M., Martin, J.-C., Pelachaud, C., Riera, M. and Wittenburg, P.: Guidelines for the Creation of NIMM Annotation Schemes. ISLE Deliverable D9.2, February 2003.

Knudsen, M. W., Bernsen, N.O., Dybkjær, L., Hansen, T., Mapelli, V., Martin, J.-C., Paulsson, N., Pelachaud, C., and Wittenburg, P.: Guidelines for the Creation of NIMM Data Resources. ISLE Deliverable D8.2, February 2003.

Knudsen, M. W., Martin, J.-C., Dybkjær, L., Ayuso, M. J. M, N., Bernsen, N. O., Carletta, J., Kita, S., Heid, U., Llisterri, J., Pelachaud, C., Poggi, I., Reithinger, N., van ElsWijk, G. and Wittenburg, P.: Survey of Multimodal Annotation Schemes and Best Practice. ISLE Deliverable D9.1, February 2002a.

Knudsen, M. W., Martin, J.-C., Dybkjær, L., Berman, S., Bernsen, N. O., Choukri, K., Heid, U., Mapelli, V., Pelachaud, C., Poggi, I., van ElsWijk, G. and Wittenburg, P.: Survey of NIMM Data Resources, Current and Future User Profiles, Markets and User Needs for NIMM Resources. ISLE Deliverable D8.1, February 2002b.

6.2 Other literature references

Bernsen, N. O.: Multimodality in language and speech systems - from theory to design support tool. In Granström, B. (Ed.): *Multimodality in Language and Speech Systems.* Dordrecht: Kluwer Academic Publishers 2002.

Brugman, H. and Wittenburg, P.: The Application of annotation models for the construction of databases and tools. In Proceedings of the IRCS Workshop on Linguistic Databases, Philadelphia, December 2001, 65-73.

Campbell, N.: Recording and Storing of Speech Data. In Proceedings of the International Workshop on Resources and Tools in Field Linguistics. pp 6.1-6.3. Las Palmas May 2002.

Wittenburg, P.: Metadata Overview and the Semantic Web. In Proceedings of the International Workshop on Resources and Tools in Field Linguistics. pp 4.1-4.14. Las Palmas May 2002.

6.3 Other website references

ATLAS: http://www.nist.gov/speech/atlas/

COCOSDA: http://www.cocosda.org/

DMCI: http://dublincore.org/

ELRA: http://www.icp.inpg.fr/ELRA/home.html

IMDI: http://www.mpi.nl/ISLE

ISO 9660: http://www.iso.org/iso/en/ISOOnline.frontpage

ISO TC37 SC4: http://www.korterm.or.kr/gita_link_doc/brochure_new.doc

LDC: http://www.ldc.upenn.edu/

OAI: http://www.openarchives.org/)

OLAC: http://www.language-archives.org/

ONOMASTICA: http://www.hltcentral.org/projects/detail.php?acronym=ONOMASTICA

SAM: http://www.icp.inpg.fr/Relator/standsam.html

SPEECHDAT: http://www.speechdat.org/

SPEECHDAT CAR: http://www.speechdat.org/SP-CAR/

SPEECON: http://www.speecon.com/

SPEX: http://www.spex.nl/

7. Appendix 1: Sample sheets

All examples in this appendix are taken from ELRA and just serve to illustrate how the ELRA agency is handling resources etc.

7.1 Sample instruction sheet

Example of an instruction sheet for telephone speech recordings:

Instructions and recommendations

Thank you for agreeing to take part in this speech collection exercise. Your voice will be one of a large collection of recorded voices which will be used to develop machines that can recognise speech. This technology can then be used to enable people to obtain services over the telephone by talking to a computer.

The task involves you making a phone call to a freephone number. Your call will be answered by a computer which will guide you through your script and also record what you say.

The recordings are completely anonymous and no-one will be 'listening-in' to you speaking.

The freephone lines are available 24 hours a day, everyday including Saturday and Sunday, however the office is only manned during the week, so it is preferable if you can call Monday to Friday. If this is not convenient you can make your call at what ever time suits you best, though it would help us if you make the call in the next few days.

The FREEPHONE number to call is:

<phone number>

Before you call please note the following points:

Please ensure that you have read and rehearsed the script carefully - do not worry if some phrases appear odd or irrational.

Please try to create a quiet environment for making the call. Background noises such as TV/radio etc., domestic equipment or children playing/crying will make it hard for you to hear what the computer is saying to you and these sorts of noises will make it difficult for the computer to record your voice properly. Please also avoid having a conversation with anyone else during your call.

Choose a time when you are unlikely to be disturbed so that you can complete the call without any interruptions or distractions.

Speak naturally - there is no 'correct' way to speak to the computer, just use your normal speaking voice (as if you were speaking to a friend).

You must follow the progress of your call on the script that you have been provided with. Following this will tell you what you should say in reply to the computer. You may find it helpful to have a pen to help you keep your place in the script and to tick off the items.

When you call the freephone number, the computer will answer and introduce itself. It will then proceed to take you through the script, stage by stage.

Each prompt from the computer will be followed by a 'tone' sound which indicates that the computer is ready to record your reply. You must wait until you have heard the tone before you answer. If you don't your speech will not be properly recorded.

You will first of all be asked to give your ID number. This is the 9-digit number which is printed at the top right-hand corner of your script. It is important that you give this number correctly as this number will be used to identify your script.

You will then be asked to give your name. Please reply with your first name only.

After you have given your ID code and name you will be asked to say the words, sentences, numbers and names that are printed on the attached script. The computer will state the appropriate item number and you should respond, after the tone, by reading the text which corresponds to that item number on your script. Please note that all items to be read are in upper case.

You will also be asked some questions for which you have to supply your own answers. These occur wherever the following statement appears on your script.

“You will now be asked a question. Please answer the question played by the system.”

The recording system has been designed to provide ample time for any speaker to read the text at his or her natural speed. Therefore there sometimes may be noticeable gaps of silence between the end of your answer and the start of the next prompt. Do not worry about such gaps, nothing has gone wrong - just hang on for a short time and the next prompt will come.

You will find it helpful to have the following pieces of information written down and beside you when you make the call. You will need this information for some of the questions.

1. Your ID (from top right-hand corner of either page of the prompt sheets)
2. The spelling of your name (please spell your first name only)
3. Your place of birth
4. Your date of birth
5. The time of day
6. A telephone number - yours or somebody else’s.

If you run into any difficulties during a call, simply ring off and try again.

Please persevere to the end of the script as you will only be paid if you complete the call

Further Information

Please note that by taking part you are consenting to the use of your recorded speech in speech technology research and product development. Thank you again for your help.

7.2 Sample prompt sheet

Example of a prompt sheet for speech recordings:

ID: 081005348

ITEM NUMBER TEXT

#1 JUST AFTER TWENTY-FIVE MINUTES TO TEN IN THE MORNING

#2 0430 9686 2602 6254

#3 THE TREES ERUPTED WITH THE SCREAMS OF A DOZEN MONKEYS

#4 LANGUAGE

#5 BECAUSE WE WANT TO SEE HOW FAR HE'LL GO THIS TIME

#6 YESTERDAY

#7 DIAL

#8 8146

#9 REMEMBER

#10 SAVE

#11 SO THE POLICE HAD TO LET ALL OF THEM GO

#12 BROUGHTON

#13 £ 29.40

#14 DOING

#15 You will now be asked a question. Please answer the question played by the system.

#16 AN OFF SPIN BOWLER WILL SOMETIMES NOT SPIN THE BALL SO MUCH

#17 THE REASONS FOR FLAGGING OUT ARE QUITE CLEAR TO MANY SHIP OWNERS

#18 You will now be asked a question. Please answer the question played by the system.

#19 THE AMOUNT OF CLEANING A PLATE REQUIRES DEPENDS UPON THE STATE IT IS IN

#20 JUST AFTER TWENTY MINUTES TO TEN AT NIGHT

#21 4

#22 You will now be asked a question. Please answer the question played by the system.

#23 SATURDAY THE TWENTY SIXTH OF OCTOBER 1996

#24 ANSWERS

#25 You will now be asked a question. Please answer the question played by the system.

#26 THEY JUMPED FROM TREE TO TREE

#27 TONY ROBERTS

#28 You will now be asked a question. Please answer the question played by the system.

#29 D E E D P O L L

#30 WHEN YOU ARE ILL YOU SHOULDN'T ACT

#31 HER AMBITION TO LEARN FRENCH OVER THE SUMMER EARNED HER NEARLY UNANIMOUS RIDICULE

#32 You will now be asked a question. Please answer the question played by the system.

#33 MERRILL LYNCH

#34 907807

#35 THE BALL IS RELEASED NEAR THE TOP OF THE ARC

#36 PLEASE FILE HER E-MAIL IN MY PENDING FOLDER.

#37 B R O U G H T O N

#38 You will now be asked a question. Please answer the question played by the system.

#39 9 7 5 3 1 2 4 6 8 0 zero naught

#40 PLOUGHSHARE

#41 SELECT ENGLISH.

#42 COWES

#43 01868 442 6962

7.3 Recording waiver (adults)

RECEIPT

Last_name :
……………………………………………………………………………………………………

First_name :
……………………………………………………………………………………………………

Street :
………………………………………………………………………………………………..……

Post_code :………………………

City : ………………………………………………...……………….

Tel. :
…………………………………………………………………………………………………….

Hereby I certify that I have received the sum of X Euro from company Y and by this I authorize Y to use the recordings of my voice/face for the purpose of research and development of new technologies.

Date, place and signature

7.4 Recording waiver (children)

PARENTAL AUTHORIZATION

Last_name : ……

First_name : ……

Street : ……………………………………………………………………………………………………….……

Post_code :………………………

City : ……………………………………………………...……………….

Tel. : …….

Hereby I authorize my child ………………………………………………….. to participate in the recordings for company Y which takes place in location Z. I testify that my child has received the sum of X euros and I authorize Y to use the recordings of my voice/face for the purpose of research and development of new technologies.

Date, place and signature

7.5 Description form – Speech

S.1. General Information

Type of resource:
□ Telephone fixed □ Telephone mobile □ Telephone IP
□ Desktop/Microphone □ Broadcast news □ Phonetic lexicon
□ Other:

Acquisition mode :
□ Acoustic □ Articulatory □ Aero-dynamic
□ Physiologic □ Other :

Speech style(s) :
□ Spontaneous □ Read □ Elicited
□ Prepared □ Prompted □ Other :
Specification (e.g. interview, casual conversation, etc.) :

Speech content: No. of items: No. of items:
□ Application words _______ □ Digit-set _______
□ Concatenated words _______ □ Isolated digits _______
□ Isolated words _______ □ Continuous sentences _______
□ Syllables _______ □ Phonetically rich sentences _______
□ VCV sequences _______ □ Phonetically balanced sentences _______
□ Yes/no questions _______ □ Other: _______

Speech setting:
□ Monologue □ Dialogue □ Multilogue

Recording scenario(s):
□ Office □ Other room □ Public place (open)
□ Public place (closed) □ Moving vehicle □ Vehicle standing still
□ Dead room □ Other :
Comments :

Microphone type:

Telephone type:

Network type:

Application:
□ Discourse analysis □ Language identification □ Speaker identification
□ Speaker verification □ Speech recognition □ Spoken dialogue systems
□ Voice control □ Other:

S.2. Speaker Specific Information

Sex and number of speakers:

□ Male Number:

□ Female Number:

□ Impostors Number:

Total number:

Age class: (indicate number)

□ Children (up to 12) □ Teenagers (12-15) □ Teenagers (16-19)

□ Adults young (20-30) □ Adults (31-45) □ Adults (46-60)

□ Elderly (over 60) □ Age unknown □ Other distribution:

Comments:

Origin:

□ Native □ Non native □ Unknown

Comments:

Geographic distribution:

Total number of regions:

Percentage per region:

Regions included:

Information included about:

□ Place of living □ Place of birth □ Place of (secondary) education

□ Dialect/accent

Comments:

Additional speaker information included:

□ Speaking/hearing impairments □ Height □ Weight

□ Smoking habits □ Trained speakers □ Education level

□ Profession

Comments:

S.3. Lexicon

Lexicon included: □ Yes □ No

Size (number of lexicon entries) :

Format:

□ ASCII □ SGML □ TEI

□ Other:

Pronunciation lexicon:

□ Available □ Not available

Transcriptions:

□ Canonical only □ Canonical + alt. pronunciation □ Automatically generated

□ Checked manually □ Generated fully manually □ Other:

Phoneme set:

□ IPA □ SAMPA □ CPA

□ Other:

S.4. Linguistic Information and Segmentation

Linguistic annotation:

□ Orthographic □ Morphological □ Phonetic

□ Syntactic □ Semantic □ Prosodic

□ Other:

Level of segmentation:

Level of annotation:

S.5. Technical Information

Signal encoding:

□ A-law □ μ-law □ Linear

□ PCM □ Other:

File format:

□ AIFF □ Wav □ Without header

□ SAM □ NIST/Sphere □ Au

□ Other:

Sampling rate:

□ 8 kHz □ 16 kHz □ 32 kHz

□ 44,1 kHz □ 48 kHz □ Other:

Quantisation:

□ 8 bit □ 16 bit □ 32 bit

□ Other:

Byte order:

□ Lo-hi (Intel) □ Hi-lo (Motorola)

Data format:

□ Signed integer □ Unsigned integer □ Floating point

□ Other:

Amount of data:

Size (Mb, Gb, etc) or duration (minutes, hours, etc):

Compression:

□ None □ Zip □ Shorten

□ Other:

Number of recording channels:

□ 1 (mono) □ 2 (stereo) □ 3

□ 4 □ 8 □ Other:

Annotation standard:

□ SAM □ SGML □ XML

□ NIST/LDC □ Other:

Sound quality measures included:

□ SNR □ Cross talk □ Clipping rate

□ Background noise □ Other:

Tools used for measuring sound quality:

S.6. Further Comments

7.6 Description form – Multimodal

M.1. General Information

Data included:

□ Audio *(see section M.6.)* □ Image *(see section M.7.)* □ Video *(see section M.8.)*

Language(s):

□ Language dependent □ Language independent

Language(s):

M.2. Recording Information – Humans

Sex and number of humans:

□ Male Number:

□ Female Number:

□ Imposters Number:

□ Synthetic Number:

Total number:

Number of humans visible in the same frame:

Age class: (indicate number)

□ Children (up to 12) □ Teenagers (12-15) □ Teenagers (16-19)

□ Adults young (20-30) □ Adults (31-45) □ Adults (46-60)

□ Elderly (over 60) □ Age unknown □ Other distribution:

Comments:

Origin:

□ Native □ Non native □ Unknown

Comments:

Geographic distribution:

Total number of regions:

Percentage per region:

Regions included:

Information included about:

□ Place of living □ Place of birth □ Place of (secondary) education

□ Dialect/accent

Comments:

Additional subject information included:

□ Speaking/hearing impairments □ Height □ Weight

□ Smoking habits □ Trained subjects □ Education level

□ Profession

Comments:

M.3. Recording Information – Resource

Human body parts visible in the resource:

□ None □ Whole body □ Head

□ Face □ Mouth □ Arms

□ Hands □ Legs □ Feet

□ Other:

Distractors visible in the resource:

□ None □ Hat □ Glasses

□ Watch □ Scarf □ Pen/Paper/Notepad

□ Microphone □ Markers □ Mobile phone

□ Other:

Interactive media visible/audible in the resource:

□ None □ Graphical screen □ Computer pen

□ Tactile screen □ Data glove □ PDA

□ Desktop □ Laptop □ Mouse

□ Bluetooth □ Other:

Annotated modalities in the resource: *(for details see section M.4.)*

□ None □ Speech □ Hand/Arm gestures

□ Gaze/Eye movements □ Facial expressions □ Lip movements

□ Head movements □ Body movements □ Hand manipulation of objects

□ Other:

Total number of annotated modalities:

Other modalities available/visible but not annotated in the resource: *(for details see section M.4.)*

□ None □ Speech □ Hand/Arm gestures

□ Gaze/Eye movements □ Facial expressions □ Lip movements

□ Head movements □ Body movements □ Hand manipulation of objects

□ Other:

Total number of modalities (not annotated):

Scene – Illumination:

□ Daylight □ Single source □ Multiple sources

□ Fix □ Variable □ Other:

Scene – Backgrounds:

□ Plain □ Complex □ Other:

Data:

Total number of sessions:

Number of poses per subject:

M.4. Modalities – Detailed Information

Facial Expressions (universal expressions of emotion):

□ Surprise □ Fear □ Disgust

□ Anger □ Happiness □ Sadness
□ Shame □ Embarrassment
□ Other:

Head movements:
□ Rotation □ Inclination forward/backward □ Inclination sideward
□ Other:

Face views:
□ Frontal □ Profile □ Other:
Total number of face views per subject:

Gestures:
Communicative:
□ Iconic □ Deictic □ Beat
□ Metaphorical □ Other:
Instructional:
□ Pointing □ Grasping □ Other:

Gaze/Eye movements:
□ Saccades □ Pursuit motion □ Convergence
□ Horizontal □ Vertical □ Other:

Hand manipulation of objects:
□ Direct manipulation □ Indirect manipulation □ Modifying objects
□ Joining/splitting objects □ Changing object position □ Other:

Body movements:
□ Upper body □ Lower body □ Both
□ Other:

M.5. Application Information

Authentication:
□ Face verification □ Speech verification □ User authentication
□ Other:

Recognition:
□ Face recognition □ Automatic speech recognition □ Automatic person recognition
□ Expression recognition □ Other:

Analysis:
□ Lip tracking □ Speech/lips correlation □ Other:

Synthesis:
□ Talking heads □ Avatars □ Humanoid agents
□ Multimedia development □ Other:

Control:
□ Voice control □ Speech assisted video □ Other:

Miscellaneous:

□ Information retrieval □ Other:

Application areas:

□ Education/Training □ Research □ Entertainment

□ Banking □ Tourism □ Other:

M.6. Technical Information – Audio

Signal encoding:

□ A-law □ μ-law □ Linear

□ PCM □ Other:

File format:

□ AIFF □ Wav □ Without header

□ SAM □ NIST/Sphere □ Au

□ Other:

Sampling rate:

□ 8 kHz □ 16 kHz □ 32 kHz

□ 44,1 kHz □ 48 kHz □ Other:

Quantisation:

□ 8 bit □ 16 bit □ 32 bit

□ Other:

Byte order:

□ Lo-hi (Intel) □ Hi-lo (Motorola)

Data format:

□ Signed integer □ Unsigned integer □ Floating point

□ Other:

Amount of data:

Size (Mb, Gb, etc) or duration (minutes, hours, etc):

Compression:

□ None □ Zip □ Shorten

□ Other:

Number of recording channels:

□ 1 (mono) □ 2 (stereo) □ 3

□ 4 □ 8 □ Other:

Annotation standard:

□ SAM □ SGML □ XML

□ NIST/LDC □ Other:

Sound quality measures included:

□ SNR □ Cross talk □ Clipping rate

□ Background noise □ Other:

Tools used for measuring sound quality:

Speech content: No. of items: No. of items:

□ Application words _______ □ Digit-set _______

□ Concatenated words _______ □ Isolated digits _______

□ Isolated words _______ □ Continuous sentences _______

□ Syllables _______ □ Phonetically rich sentences _______

□ VCV sequences _______ □ Phonetically balanced sentences _______

□ Yes/no questions _______ □ Other: _______

M.7. Technical Information – Image

Resolution in pixels:

Colour components:

□ RGB □ CMYK □ 4:2:2

□ Other:

Colour depth:

□ 8 bits □ 16 bits □ 24 bits

□ Other:

File format:

□ JPG □ GIF □ TIFF

□ BMP □ EPS □ CIF

□ PPM □ Other:

Amount of data:

Size (Mb, Gb, etc):

Duration (minutes, hours, etc):

Compression:

□ None □ Zip □ Other:

Compression ratio:

M.8. Technical Information – Video

□ Synchronized audio *(see section M.6.)*

Resolution in pixels:

Colour components:

□ RGB □ CMYK □ 4:2:2

□ Other:

Color depth:

□ 8 bits □ 16 bits □ 24 bits

□ Other:

Frame rate:

□ <25 frames/sec. □ 25 frames/sec. □ 30 frames/sec.

□ 50 frames/sec. □ 60 frames/sec. □ Other:

File format:

□ MOV □ AVI □ MPEG

□ QuickTime □ SGI □ Other:

Amount of data:

Size (Mb, Gb, etc):

Duration (minutes, hours, etc):

Compression:

□ None □ Zip □ Other:

Compression ratio:

M.9. Technical Information – Modelling

Models:

□ 2D □ 3D □ Other:

File formats:

□ VRML □ Other:

Algorithms used:

M.10. Further Comments

7.7 Contract for distribution

LANGUAGE RESOURCES DISTRIBUTION AGREEMENT

BETWEEN

"…………………………………….."

And

"……………………………………."

This AGREEMENT is made by and between:

"…………………………..", (hereinafter called PROVIDER), having its principal place of business at:
……………………………………...

AND

"…………………………..", (hereinafter called DISTRIBUTOR), having its principal place of business at:
……………………………………...

Terms and conditions

PROVIDER certifies that he is the rightful holder of the Languages Resources described in Exhibit A.

PROVIDER grants DISTRIBUTOR, who accepts, the non-exclusive right to distribute the Language Resources described in Exhibit A. "Distribution" shall mean that PROVIDER enables DISTRIBUTOR to market the Language Resources according to DISTRIBUTOR's marketing, distribution and commercialisation policies.

PROVIDER authorizes DISTRIBUTOR to grant USER Licenses for the use of the Language Resources to any legal entity. DISTRIBUTOR shall impose the relevant obligations of this AGREEMENT on such entity.

The Language Resources may be duplicated by DISTRIBUTOR as indicated in Exhibit B. DISTRIBUTOR is also authorized to reproduce, in whole or in part, and to modify the Language Resources, as well as the accompanying DOCUMENTATION and MANUAL for the purposes of distribution.

DISTRIBUTOR agrees to pay PROVIDER a compensation. The mode of payment and schedule of payments are incorporated in Exhibit C.

DISTRIBUTOR shall give appropriate references to PROVIDER in scholarly literature when the Language Resources are mentioned. DISTRIBUTOR shall not use the name of PROVIDER in any publication in any manner that would imply an endorsement of DISTRIBUTOR or any product or service offered by DISTRIBUTOR.

PROVIDER gives no warranty for merchantibility and/or fitness for a particular purpose of the Language Resources.

DISTRIBUTOR gives no warranty for the commercial success of its marketing efforts.

Both parties exclude all liability of whatsoever nature for direct, consequential or indirect loss or damage suffered by the other, in connection with the distribution of Language Resources.

Neither party shall be held responsible for any delay or failure in performance caused by « force majeure » or other causes beyond the parties' control and without the parties' fault or negligence. Should such event occur, all obligations in this AGREEMENT should be sustained throughout the duration of the event.

The entire AGREEMENT is composed of the 10 articles herein together with Exhibits A, B, and C thereafter.

In witness whereof, intending to be bound, the parties hereto have executed this AGREEMENT by their duly authorized officers:

AUTHORIZED BINDING SIGNATURES :

_________________	_________________
On behalf of Provider	On behalf of Distributor
Name:	Name: CEO
Title:	Title: Managing Director
Date:	Date:

EXHIBITS

EXHIBIT A: Language resources description:

EXHIBIT B: MEANS OF DATA DELIVERY:

Means of delivery:

EXHIBIT C : PAYMENT SCHEDULE:

Sales are notified to Provider every semester (end of December and end of June) in writing.
Payments, as defined below, shall be paid within thirty days after the receipt of invoices, by transfer of the sum concerned to the bank account number specified on respective invoices. The said amounts are exclusive of value-added tax.