

Chapter 21

The Dutch-Flemish HLT Agency: Managing the Lifecycle of STEVIN's Language Resources

Remco van Veenendaal, Laura van Eerten, Catia Cucchiarini,
and Peter Spyns

21.1 Introduction

The development and availability of human language technologies is considered crucial for a language to be able to survive in the information society. Since Dutch is a so-called mid-sized language [11, 12] with a relatively small market, companies tend to be reluctant to invest in the development of resources for such a language. If they do, the resulting data are not always made available to researchers or other companies at affordable prices. Hence governmental support is required.

The Dutch Language Union (NTU – Nederlandse Taalunie¹), an intergovernmental organisation established by Belgium and the Netherlands, has stated as one of its priorities the promotion of the development of Human Language Technology (HLT) for the Dutch language [3, 6]. In co-operation with the relevant ministries and organisations in Belgium and the Netherlands, the NTU set up a number of initiatives (cf. Chap. 2, page 21). Two of these are particularly significant: the STEVIN research programme [13] and the Dutch-Flemish Human Language Technology Agency (HLT Agency, TST-Centrale in Dutch) [2]. In addition, these two initiatives are related as it was stipulated that the language resources developed within the STEVIN programme would later be handed over to the HLT Agency for subsequent management, maintenance and distribution.

¹See http://taalnieversum.org/taalunie/how_can_we_help_you_/index.php.

R. van Veenendaal (✉) · L. van Eerten
Dutch-Flemish HLT Agency (TST-Centrale), Institute for Dutch Lexicology, Matthias de Vrieshof 2-3, 2311 BZ Leiden, The Netherlands
e-mail: remco.vanveenendaal@inl.nl; laura.vaneerten@inl.nl

C. Cucchiarini · P. Spyns
The Nederlandse Taalunie, Lange Voorhout 19, 2514 EB Den Haag, The Netherlands
e-mail: ccucchiarini@taalunie.org; pspyns@taalunie.org

The establishment of a central repository for managing LRs prevents LRs developed with public money from becoming obsolete and therefore useless. Resources that are not maintained quickly lose value. In the past, official bodies such as ministries and research organisations used to only finance the development of LRs and did not have a clear policy on what should happen to these materials once the projects had been completed. Universities rarely have the resources to do maintenance work on completed projects, and knowledge is sometimes lost when experts switch jobs. It was against this background that the idea of having one central repository for digital language resources in the Dutch language area was conceived.

Having one organisation that is responsible for managing LR lifecycles creates higher visibility and better accessibility. Having a local one-stop-shop for Dutch LRs leads to more re-use of these LRs. Synergistic use of manpower and means is efficient and cost-reducing, compared to, for example, having several (smaller) organisations that only take care of certain aspects of LR lifecycles instead of their complete lifecycles. Furthermore, combining resources and bringing together different kinds of expertise creates surplus value. This can result in improved versions of datasets and new insights into potential use(r)s of LRs.

In this chapter, the HLT Agency is presented mainly from the perspective of its contribution to the STEVIN programme. In Sect. 21.2, the organisational set-up of the HLT Agency is described. Section 21.3 explains how the HLT Agency manages the lifecycle of STEVIN results. In Sect. 21.4, target groups and users are presented. Section 21.5, discusses new challenges and possible directions for the HLT Agency after the completion of the STEVIN programme. Section 21.6 concludes the chapter and points to future perspectives.

21.2 The Flemish-Dutch HLT Agency

The HLT Agency is an initiative of the NTU, from which it currently receives an annual subsidy of 450,000 euros. Additional financing comes mostly from projects, support and licence fees. The HLT Agency can be considered a non-profit, government-funded initiative. It is currently hosted at the Institute for Dutch Lexicology (Instituut voor Nederlandse Lexicologie, INL), a Dutch-Flemish organisation with offices in the Netherlands (Leiden) and Belgium (Antwerp).

The mission given by the NTU to the HLT Agency was to maintain, manage and distribute LRs for Dutch, in particular those owned by the NTU, which include the STEVIN results. This also implies clearing IPR issues with the suppliers of the LR at the in-take (acquisition licences – cf. Sect. 21.3.1) and safe-guarding the interests of the owners of the LR towards users of the resources (distribution licences – cf. Sect. 21.3.4). STEVIN is not the only source of LRs that are hosted by the HLT Agency. Aside from STEVIN, the HLT Agency tries to collect high-quality digital Dutch LRs from funding institutions such as the NTU, the INL, funding agencies like the NWO and third parties such as (individual) researchers,

universities and some other organisations and foundations. In fact, the first LRs the HLT Agency acquired came from previous government-funded projects (the Spoken Dutch Corpus [10], the Referentiebestand Nederlands,² both now property of the NTU, and several lexical resources of the INL).

The HLT Agency's core team consists of a project manager, three linguists, and a co-worker who takes care of all distribution procedures. Each linguist is responsible for a particular group of LRs. Additionally, one of the linguists doubles as a communications officer, another as manager of the licences. Most employees work part-time. In addition to the core team, one programmer, two computational linguists and one system administrator contribute to the HLT Agency part-time. They are regular employees of the INL and support the HLT Agency with technical work on LRs, such as updates or tailor-made versions of LRs. .

Other repositories for managing LRs exist (cf. [8, p. 39] for an overview). Two examples are DANS (in the Netherlands) and ELDA (in Europe):

- Data Archiving and Networked Services (DANS),³ an institute of the Royal Netherlands Academy of Arts and Sciences (KNAW)⁴ and the Netherlands Organisation for Scientific Research (NWO),⁵ promotes sustained access to digital research data. For this purpose, DANS encourages researchers to archive and reuse data in a sustained manner, e.g., through an online archiving system.
- The European Language resources Distribution Agency (ELDA)⁶ is the operational body of the European Language Resources Association (ELRA)⁷ and was set up to identify, classify, collect, validate and produce language resources. ELDA is also involved in HLT evaluation campaigns.

Whereas DANS focuses on all research areas and ELDA on LRs in many languages, it is the HLT Agency's specific mission to take care of Dutch digital LRs in order to strengthen the position of the Dutch language in the information society. The HLT Agency therefore focuses on digital Dutch LRs (as opposed to ELDA) and needs to ensure that LRs are not only made available (as opposed to DANS), but also kept up-to-date and usable. This means the HLT Agency takes care of the management of the entire lifecycle of LRs, including maintenance and support. Creating new resources and performing evaluation campaigns currently do not belong to the mission of the HLT Agency (as opposed to ELDA).

²See http://www.tst-centrale.org/images/stories/producten/documentatie/rbn_documentatie_nl.pdf

³See <http://www.dans.knaw.nl>

⁴See <http://www.knaw.nl>

⁵See <http://www.nwo.nl>

⁶See <http://www.elda.org>

⁷See <http://www.elra.info>

21.3 Managing the Lifecycle of STEVIN Results

The HLT Agency distinguishes five different phases in the lifecycle management process: acquisition, management, maintenance, distribution and support. These phases are described in the following subsections.

21.3.1 Acquisition

From the very start of the STEVIN programme, the funding partners agreed that the NTU would become the owner of the LRs developed within the STEVIN projects and that these would be transferred to the HLT Agency. The rationale behind this decision was to ensure optimal LR accessibility (see below). STEVIN project proposals included a description of the potential future use of the resulting resources and of the contribution to the overall STEVIN aims (cf. Chap. 2, page 25). A positive review and subsequent funding of the project implied that the resources were worthwhile to be maintained and subsequently distributed. Therefore the process of acquisition was to a large extent fixed. The HLT Agency focused on settling issues concerning intellectual property rights (cf. Sect. 21.3.1.1) and checking the quality of the LRs delivered by the STEVIN projects (cf. Sect. 21.3.1.2).

21.3.1.1 Intellectual Property Rights

In general, the intellectual property rights (IPR) of the LRs developed and completed in the STEVIN programme (the foreground knowledge) were transferred to the NTU. However, if pre-existing software tools or data (background knowledge) were used in the projects, the rights on the background knowledge would remain with the original IPR-holder. If existing open source was re-used and improved, the resulting STEVIN LR would also become open source (as was the case for e.g., the STEVINcanPRAAT project – cf. Chap. 5, page 79).

Transferring all rights to one organisation, c.q. the NTU, has several practical advantages. IPR issues such as setting up licences and price negotiations can be dealt with more efficiently, both in terms of money and time; the more IPR holders, the more rights and responsibility issues. It is also a guarantee that IPR issues are dealt with in a legally sound manner, which in turn leads to considerably fewer restrictions on LR availability.

With regard to the data themselves, it was primarily the responsibility of the STEVIN projects to settle IPR issues such as copyright on texts and usage of speech data. The project proposals had to include a section on how IPR issues would be taken care of. The role of the HLT Agency was to provide assistance to projects during the process. As a result, the HLT Agency was involved in the process of settling IPR issues from an early stage.

Altogether, corpus creation projects like SoNaR (cf. Chap. 13, page 219) and DPC (cf. Chap. 11, page 185) have resulted in more than 150 signed data acquisition licences. Tailor-made versions of these licences were made available after discussing requests from data providers with the project team, a lawyer and the NTU. One example of a tailor-made version is a licence for publishers who are willing to provide data for research purposes only (while the standard licence also includes the use of the data for commercial purposes). The HLT Agency⁸ acted as signing party for acquisition licences on behalf of the NTU.

21.3.1.2 Evaluation and Validation

STEVIN requested projects to have their deliverables externally evaluated. Evaluation is necessary to gauge the LR's quality and potential for re-use.

After completion of the projects, the evaluation reports were handed over to the HLT Agency, together with the LRs. The HLT Agency monitors the value indicators for the purpose of prioritising future maintenance work. In the case of e.g. the IRME (cf. Chap. 12, page 201), JASMIN-CGN (cf. Chap. 3, page 43) and Cornetto (cf. Chap. 10, page 165) projects, conducting an early external evaluation resulted in the delivery of improved project results at the end of the project.

The HLT Agency's active contribution to the validation process is limited to technical checks. All data are validated (e.g., against XML schemas) and the quality and completeness of the documentation is thoroughly checked. In the case of software the binaries are tested on the supported platforms, using the accompanying documentation to create test cases. Source codes are compiled to (object codes and linked into) testable binaries. If the data validation produces significant errors, software does not execute or does not work as expected (according to the documentation), it is primarily the LR provider's task to fix any problems and resubmit the fixed LR to the HLT Agency.⁹

21.3.2 Management

The LRs delivered by STEVIN projects are stored and backed up on servers hosted and maintained by the INL, the HLT Agency's hosting institute. Where needed, LRs are stored in version control software like Subversion.¹⁰ Our archive and "production line system" [4] servers currently contain 1.5 terabyte and over 60 LRs.

⁸In fact, the INL signs the licences as the INL is the HLT Agency's legal entity.

⁹This practice is in line with the Data Seal of Approval guidelines (see <http://www.datasealofapproval.org>), adopted by the HLT Agency.

¹⁰See <http://subversion.apache.org/>. One reason for setting up Subversion for an LR is collaborative maintenance work.

21.3.3 *Maintenance*

The STEVIN LRs, including all accompanying deliverables (project proposals, reports and documentation), are stored and backed up by the HLT Agency in their original form. A separate distribution version is prepared, consisting for example of the LR and relevant user documentation. If the LR provider agrees, the evaluation report is included. Periodically, the HLT Agency checks if LRs need maintenance, e.g. for the purpose of standardisation or when they risk disuse due to incompatibility with new operating systems. Also actual use and peer reviews (user feedback) of the LRs give indications of whether or not LRs need maintenance. The HLT Agency distinguishes between minor and major maintenance.

Minor Maintenance The goal of minor maintenance is to keep resources usable, which means fixing critical bugs, updating manuals and documentation, upgrading formats to newer versions of the standard(s) used, etc. Minor maintenance is done by the HLT Agency itself. Periodically, the HLT Agency checks if LRs require minor maintenance and starts the work after having consulted the owner/supplier of the LR. Feedback from users is included in these maintenance checks. The result of minor maintenance is usually a patch or update of an LR. News on any updated versions is published, so that users can request an update for free.

Major Maintenance Major maintenance consists in significantly improving or expanding a resource. Therefore, major maintenance usually requires additional funding and cooperation with the developers and external experts. Information and advice on which LRs should be improved or expanded can be gathered from the various advisory committees that assist the NTU and the HTL Agency and from user feedback collected by the service desk. Major maintenance work usually results in a new version of an LR, rather than a patch or update. News on any new versions is published, for which all users must accept a (new) licence.

Minor and Major Maintenance in Practice Below we present some examples of maintenance on to the STEVIN IRME (cf. Chap. 12, page 201) and Cornetto (cf. Chap. 10, page 165) projects:

- IRME
 - In the final stage, but before the end of the project, version 1.0 of the DuELME resource (Dutch Electronic Lexicon of Multiword Expressions) was updated by the project members at Utrecht University after they had received the external validation report. The resulting version 1.1 was made available to the HLT Agency for further distribution.
 - The HLT Agency improved the Web interface for the lexicon by making minor adjustments in functionality and display to the search tool and by optimising the MacOSX-based DuELME web interface for Windows (minor maintenance).

- Utrecht University and the HLT Agency converted the DuELME lexicon to LMF (Lexical Markup Framework) within the CLARIN-NL¹¹ project DuELME-LMF. This became version 2.0 of the resource (major maintenance). This version was also made available in the CLARIN infrastructure.
- Cornetto
 - In response to user requests, intermediate versions of the Cornetto database (Combinatorial and Relational Network as Tool kit for Dutch Language Technology; a lexical semantic database for Dutch) were made available during the project. The HLT Agency took care of the licences and the project team, led by the Free University Amsterdam (VUA), distributed and supported the database.
 - VUA and the HLT Agency improved version 1.0 of the Cornetto database to versions 1.2, 1.21 and 1.3 (minor maintenance).
 - Currently, VUA is significantly improving the Cornetto database. For example, sentiment values and text corpus references are being added for each word meaning. This work will result in version 2.0, to be released by the HLT Agency early 2012.
 - VUA and the HLT Agency will work on a further improved version of the Cornetto database in the CLARIN-NL Cornetto-LMF-RDF project (major maintenance). As a result, Cornetto will also become available within the CLARIN infrastructure, in LMF, RDF and SKOS formats.

21.3.4 Distribution

The HLT Agency makes the LRs available for users through a web shop¹²: users order an LR from the web shop and receive the LR after accepting an end user agreement. Most of the LRs are available as a downloadable file or through a web interface. Larger LRs are distributed off-line on DVDs or a hard disk. In the case of off-line distribution, the HLT Agency charges a small handling and shipping fee.¹³

The terms and conditions for the use of LRs made available by the HLT Agency are defined in (distribution) licence agreements. These agreements were written with the specific goals of the HLT Agency in mind: they have to stimulate the reuse of the LRs, but also support the idea of a central location where LRs are made and kept available. Feedback from users, stakeholders and legal experts has helped us improve and standardise the licences over the years. Other distribution centres, with different goals, apply other terms and conditions – e.g., [5].

¹¹See <http://www.clarin.nl>

¹²A new HLT Agency web site with web shop has been launched in September 2011.

¹³Currently 50 euros are charged for (a set of one or more) DVDs and 100 euros for a hard disk.

In order to strengthen the position of the Dutch language in today's information society, it is necessary to stimulate the use of Dutch in research, education and commercial end user applications. This is considered more important than financial return on investment. It implies that the HLT Agency, also due to the relatively limited size of the Dutch language area, is not supposed to become self-sustainable (as opposed to e.g., ELDA) – cf. Sect. 21.2. In short, three licensing schemes are available:

- Single licensing (non-commercial or open-source licence only)
- Dual licensing (non-commercial and commercial licence)
- Dual licensing (open-source or commercial licence)

Open-source licences have to be commonly-accepted and non-viral open-source licences. Various possibilities and variants exist [9]. Our non-commercial licences are for non-commercial use by non-commercial organisations. There is no licence fee attached to non-commercial licences (apart from incidental exceptions due to third-party rights). They do contain a right of first refusal, prohibiting the distribution of derivative works. This right of first refusal supports the idea of a “one stop shop” for LRs.¹⁴

The commercial licences have a reasonable licence fee and do not limit the distribution of derivative works. The main reason for having a licence fee for commercial licences is that we do not want to disturb the existing commercial market for Dutch LRs, however small, by making our (government-funded) LRs available for free. The fee can be settled with a one-time lump sum or with a royalty scheme over a period of time. The former has the benefit of a reduced administrative burden (one single payment once and for all), while the latter requires less money to be put on the table upfront but implies an administrative follow-up process on potential revenues.

Within the framework of the STEVIN programme, a Pricing Committee was set up to advise the NTU and HLT Agency on LR pricing and licensing matters. The nine members of this committee come from Dutch and Flemish companies, funding bodies, government organisations and research institutes and were selected because of their expertise in business, open innovation, valorisation and technology transfer.

For LRs that are acquired outside of the STEVIN programme, certain procedures are different. The transfer of IPR to the NTU, for example, is not obligatory; the rights to the LRs remain with the developers. Furthermore, open source is a standard licence model for the HLT Agency. This was not the case with STEVIN. Evaluation and validation are conducted by the HLT Agency, unless they are already part of the LR's production process.

¹⁴Note that STEVIN project consortium members “automatically” receive a distribution licence when they, as suppliers, transfer their foreground knowledge to the NTU. This allows them to continue their work using “their” project results.

21.3.5 Services

The support that the HLT Agency provides to the LRs is based on knowledge management. Knowledge management is important for at least two reasons. Firstly, the availability of the knowledge does not stay limited to the availability of the expert(s) and secondly, once collected, the knowledge can easily be used, shared, kept up-to-date and expanded. The HLT Agency ensures that LRs and all knowledge about them are made and kept available.

21.3.5.1 Sources of Knowledge

For the HLT Agency there are three primary sources of knowledge: knowledge from external experts, knowledge collected by the HLT Agency while working with or maintaining an LR and knowledge gained by the service desk through question answering.

The first source of knowledge is made available to the HLT Agency in the form of documentation and is also the result of meeting with the project team at the end of projects. Most LRs come with accompanying user and technical documentation. In addition, a considerable amount of information can usually still be obtained from the STEVIN project that created the LR, e.g. in the form of progress reports or a project wiki. The HLT Agency asks the project teams to make this information available as an additional valuable source of background knowledge. When a new LR is supplied to the HLT Agency, a knowledge transfer meeting is held with the project team. In some cases we ask the experts to explain in detail how certain parts of the LR came into existence. For example, we interviewed the lexicon expert of the Spoken Dutch Corpus project and recreated the workflow for deriving the accompanying lexicon from the corpus, which would not have been possible on the basis of the documentation alone.

Secondly, the HLT Agency creates knowledge about LRs while using and maintaining the LR. Often the user manuals of software resources created by research projects do not provide a detailed description of all possible functions. Some functionalities may not be documented at all, or they are hard to find in the user interface. Studying user manuals and software has already resulted in additional knowledge and several improved user manuals and user interfaces. Besides, a lot of knowledge is gained while maintaining LRs: working on new versions greatly improves our understanding and knowledge of LRs.

The third main source of knowledge for the HLT Agency is the question-answering provided by the service desk. The service desk is more than simply a help desk, because, a.o., it processes orders and grants access to LRs. Answers to questions are also stored and made available for reuse in case similar questions are asked. The HLT Agency has agreements with the providers or external experts regarding question-answering: when questions require knowledge that the service desk does not (yet) have, the question is forwarded to the expert. The answer provided by this expert is forwarded (with acknowledgements) and stored by the

service desk. By following this procedure, the expert does not have to come up with the same answer to the same question over and over again and the HLT Agency keeps expanding its knowledge reservoir. The service desk thus acts as a filter, reducing the amount of repetitive questions to be answered by the experts, while they are actually given credit when an answer is reused.

21.3.5.2 An Integrated Knowledge Management Cycle

Knowledge management activities start when LRs are being created: formats and standards are discussed with project teams and intermediate versions of LRs are distributed. A crucial phase in knowledge management is the moment of transition: when LRs are handed over to the HLT Agency, as much (finalised) information as possible is collected and a knowledge transfer meeting with the project team is requested. The resulting knowledge is stored in e.g. wikis (for collaborative, online work) and in documents on our servers (for finalised information). The new LR is added to the service desk and web shop and this entire process is tracked in a workflow system. Knowledge management does not end here: while managing the lifecycle of the LRs, personal and documented knowledge is updated and any new knowledge is added, for example generated in the process of LR maintenance, or resulting from answering questions through our service desk. The HLT Agency also keeps an overview of who uses the LRs for what purposes, which supports marketing efforts and helps to bring users together.

21.3.5.3 User Support

General HLT support is provided on request. Depending on the amount of work required, the HLT Agency either offers this service for free, or charges a small fee, or applies an hourly rate. Examples of this type of support are: (a) helping researchers choose appropriate standards for their data collection, (b) connecting users to organisations which can provide in their specific needs, and (c) automatically tagging data sets for others (who are not willing or able to install and use the required tools). Also other actions, useful for users and suppliers, are undertaken. E.g., the Dutch Parallel Corpus was the first LR to receive an ISBN/EAN number issued by the INL, which will facilitate referencing and citation and will improve the LR's visibility. The idea is to provide every (STEVIN) corpus with an ISBN/EAN number.

21.4 Target Groups and Users

Researchers from various disciplines turn to the HLT Agency to access all sorts of LRs, such as general, socio-, computational and forensic linguistics, translation studies, social studies, cognition studies, historical and bible studies, communication

and information studies, and Dutch studies from all over the world. Before the HLT Agency existed, researchers often had to collect their own LRs before being able to start their research proper. The advantages of this new approach in which LRs are made publicly available for researchers cannot be overestimated. For instance, since researchers do not need to allocate time and money for data collection, they can start their investigations earlier and devote more time to research. In addition, they can base their investigations on data collections that are officially documented and traceable. This is important for reviewing and replication purposes and is in line with new trends favouring open access.

Teachers and students can also access LRs for educational purposes. Frequency lists were used as a starting point in second language education or implemented in educational applications for specific groups, such as dyslectics. Audio has been used in e.g., educational games and quizzes.

Small and medium enterprises (SMEs) are another important target group for the HLT Agency. SMEs are often willing to develop useful HLT applications, but they are not always able to bear the costs of developing the LRs that are required for such applications. The availability of LRs at affordable prices through the HLT Agency lowers cost barriers and offers a viable solution. Take for example a small company that provides speech solutions for a specific user group like people with reading difficulties. The HLT Agency can offer reference lexicons, or a part of a lexicon, at a reduced price, for improving the company's speech synthesis system. The HLT Agency can also provide support or advise, based on knowledge of the LRs.

In addition to these specific target groups, a wide variety of users turn to the HLT Agency for LRs, such as lawyers, language amateurs and even artists. Examples of their use of LRs are the use of a speech corpus in a court case (where a telephone recording had to be linked to a certain person and a Dutch language model had to be constructed), the use of lexical data by crossword enthusiasts, a Dutch family abroad who wanted to teach their children Dutch, and the work on an art object incorporating speech from the Spoken Dutch Corpus.

21.5 Challenges Beyond STEVIN

With the end of the STEVIN programme, the steady and guaranteed in-flow of new LRs comes to an end. The fact that also non STEVIN projects are handing over their LRs to the HLT Agency illustrates the importance of the HLT Agency for the HLT field in Flanders and the Netherlands. Nevertheless, for the future, the NTU (as principal and funder) and the HLT Agency (as agent) have to rethink and adapt their current policies and procedures to manage the LR lifecycle, in particular regarding:

- Criteria to select which LRs are to be acquired;
- A rationale to determine the most efficient and effective manner to make LRs available;

- Guidelines to determine when LRs need which form of maintenance;
- Procedures to establish whether and which new LRs are required for Dutch;
- Strategies to raise awareness of LR availability and potential, also for companies;
- Ways to organise “knowledge platforms” and communities of practice centered around specific LRs.

One dominating element in the overall policy remains the continuous support for the Dutch language in general. Hence, the NTU may choose to favour the support (or development) of LRs that do not have a high commercial value or a large number of potential users, but which might present a high “value” for specific target group(s) of the NTU. Such LRs would simply not become available for reuse if organisations like the HLT Agency did not accept them. This seemingly resembles the Long Tail strategy: “selling a large number of unique items with relatively small quantities sold of each” [1]. However, one should not forget that this only applies for the storage and distribution aspects (via the web store) in the LR lifecycle, which are, relatively speaking, straightforward and low cost activities. Clearing the IPR, performing maintenance, and managing and expanding knowledge concerning an LR are complex and time intensive (and thus costly) activities. Hence, a judicious choice must be made about which LRs are worth spending the relatively scarce time and (public) money on (they constitute the “short tail”). The experiences gained so far and the available external expertise (e.g., the Pricing Committee – cf. Sect. 21.3.4) the HLT Agency can tap in, provide a solid basis to successfully tackle these challenges. The fact that the new South African National HLT Resource Management Agency [7] preferred a collaboration with the HLT Agency illustrates the soundness of the HLT Agency’s operating model and the professionalism of its collaborators.

Another item to consider is how the HLT Agency will position itself with respect to emerging networks, such as CLARIN and META-NET, which also aim at taking care of (parts of) the LR lifecycle. For this purpose, the NTU has become a member of the CLARIN ERIC.¹⁵ The HLT Agency already is an active partner in CLARIN-NL, the Dutch national branch of CLARIN, and will integrate into the CLARIN infrastructure as many digital Dutch LRs as possible. At the time of writing, the CLARIN-ERIC just started so that the real challenge here has only begun.

21.6 Conclusions and Future Perspectives

Thanks to STEVIN, the HLT Agency has become a linchpin of the Dutch-Flemish HLT community. Since its inception in 2004, the HLT Agency has gradually gained recognition in the HLT community in the Netherlands, Flanders and abroad. The idea of a central repository for (digital Dutch) LRs is widely supported and has

¹⁵The CLARIN-ERIC is the permanent management structure governing the CLARIN-network.

been taken up internationally. The STEVIN resources are important building blocks for the digital Dutch language infrastructure. While the STEVIN programme comes to an end in 2012, the HLT Agency will continue to act as a manager, maintainer, distributor and service desk for these and other LRs. After 7 years of mainly accompanying programmes and projects that produce LRs, the time has come to focus on the use and valorisation of their results.

In addition to STEVIN, the NTU and the INL, other parties are depositing their LRs at the HLT Agency. The sustainability of LRs is supported by adopting a clear licensing, pricing and IPR policy, maintaining the LRs, actively managing knowledge about the LRs and providing a service desk for question-answering. Although the policy and procedures adopted are subject to change over time, the goals of making and keeping digital Dutch LRs available to strengthen the position of the Dutch language in today's information society will be pursued in the future too.

In short, the HLT Agency will ensure that digital Dutch LRs, especially those derived from the STEVIN research programme, will continue to have their lifecycles properly managed and will be optimally available for research, education and commercial purposes. As of 2013, the HLT Agency is no longer hosted by the INL, but integrated in the NTU. New contact details are www.tst-centrale.org and servicedesk@hlt-agency.org.

Acknowledgements We thank the three anonymous reviewers, Anna Aalstein and Boukje Verheij for their valuable comments on earlier versions of this text.

Open Access. This chapter is distributed under the terms of the Creative Commons Attribution Noncommercial License, which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

References

1. Anderson, C.: *The Long Tail: Why the Future of Business is Selling Less of More*. Hyperion, New York (2006)
2. Beeken, J.C., van der Kamp, P.: The centre for Dutch language and speech technology (TST Centre). In: *Proceedings of the 4th International Conference on Language Resources and Evaluation (LREC04)*, Lisbon, pp. 555–558 (2004)
3. Binnenpoorte, D., Cucchiarini, C., D'Halleweyn, E., Sturm, J., de Vriend, F.: Towards a roadmap for human language technologies: the Dutch-Flemish experience. In: *Proceedings of the 3th International Conference on Language Resources and Evaluation (LREC02)*, La Valletta (2002)
4. Boekestein, M., Depoorter, G., van Veenendaal, R.: Functioning of the centre for Dutch language and speech technology. In: *Proceedings of the 5th International Conference on Language Resources and Evaluation (LREC06)*, La Valletta, pp. 2303–2306 (2006)
5. Choukri, K., Piperidis, S., Tsiavos, P., Weitzmann, J-H.: META-SHARE: Licenses, Legal, IPR and Licensing issues. META-NET Deliverable D6.1.1 (2011)

6. Cucchiari, C., Daelemans, W., Strik, H.: Strengthening the Dutch language and speech technology infrastructure. In: Notes from the Cocosda Workshop 2001, Aalborg, pp. 110–113 (2001)
7. Grover, A.S., Nieman, A., van Huyssteen G., Roux, J.: Aspects of a legal framework for language resource management. In: Proceedings of the 8th International Conference on Language Resources and Evaluation (LREC12), Istanbul, pp. 1035–1039 (2012)
8. Mariani, J., Choukri, K., Piperidis, S.: META-SHARE: Constitution, Business Model, Business Plan. META-NET Deliverable D6.3.1 (2001)
9. Oksanen, V., Lindén, K., Westerlund, H.: Laundry symbols and license management: practical considerations for the distribution of LRs based on experiences from CLARIN. In: Proceedings of the 7th International Conference on Language Resources and Evaluation (LREC10), La Valletta (2010)
10. Oostdijk, N.: Chap. The design of the spoken Dutch corpus. *New Frontiers of Corpus Research*, pp. 105–112. Rodopi, Amsterdam (2002)
11. Pogson, G.: Language technology for a mid-sized language, part I. *Multiling. Comput. Technol.* **16**(6), 43–48 (2005a)
12. Pogson, G.: Language technology for a mid-sized language, part II. *Multiling. Comput. Technol.* **16**(7), 29–34 (2005b)
13. Spyns, P., D’Halleweyn, E., Cucchiari, C.: The Dutch-Flemish comprehensive approach to HLT stimulation and innovation: STEVIN, HLT Agency and beyond. In: Proceedings of the 6th International Conference on Language Resources and Evaluation (LREC08), La Valletta, pp. 1511–1517 (2008)