A decade of HLT Agency activities in the Low Countries: from resource maintenance (BLARK) to service offerings (BLAISE)

P. Spyns[1,2] *, R. van Veenendaal[1]
[1] Nederlandse Taalunie – TST-Centrale
Lange Voorhout 19, 2514 EB Den Haag, Nederland
rvanveenendaal@taalunie.org; http://taalunieversum.org
[2] Vlaamse overheid – Departement Economie, Wetenschap en Innovatie
Koning Albert II-laan 35, bus 10, 1030 Brussel, België
Peter.Spyns@ewi.vlaanderen.be; http://www.ewi-vlaanderen.be

Abstract

In this paper we report on the Flemish-Dutch Agency for Human Language Technologies (HLT Agency or TST-Centrale in Dutch) in the Low Countries. We present its activities in its first decade of existence. The main goal of the HLT Agency is to ensure the sustainability of linguistic resources for Dutch. 10 years after its inception, the HLT Agency faces new challenges and opportunities. An important contextual factor is the rise of the infrastructure networks and proliferation of resource centres. We summarise some lessons learnt and we propose as future work to define for Dutch (which by extension can apply to any national language) a set of Basic LAnguage Infrastructure SErvices (BLAISE). As a conclusion, we state that the HLT Agency, also by its peculiar institutional status, has fulfilled and still is fulfilling an important role in maintaining Dutch as a fully fledged digital functional language.

Keywords: linguistic resources infrastructure, human language technology for Dutch, HLT Agency

1. Introduction

Languages do not confine themselves inside the boundaries of a (single) state. In 1980 the Belgian[1] and Dutch governments signed a treaty to cooperate in promoting and strengthening the position of the Dutch language and created the Dutch Language Union (Nederlandse Taalunie - NTU). They gave up a part of their autonomy and decided to conduct – to a certain degree – a joint language policy. This unique kind of cooperation has many advantages: duplication of efforts can be avoided, expertise can be shared and funds pooled.

Since long, the NTU has taken a serious interest in digital language resources and human language technologies (HLT), because they are crucial for a language to survive in the digital society. In 1999, the Dutch and Flemish governments decided to collaborate on HLT for Dutch (HLTD) and set up an HLT Platform (Beeken et al., 2000). The HLT Platform organised a number of activities, of which one eventually resulted in a stimulation programme for HLTD. It was called STEVIN and was worth more than 11.4 million euros (Spyns & D'Halleweyn 2012, Spyns and Odijk 2013, Spyns & D'Halleweyn 2013). Another very important result was the delivery of the *Blueprint for management, maintenance and distribution of digital materials developed with public money (Blueprint)* (van Sterkenburg et al. 2002) that led to the creation in September 2003 of the Dutch-Flemish Human Language Technology Agency (HLT Agency[2]) (Beeken & van der Kamp 2004), which became operational since June 2004. With both initiatives, the NTU wanted to create a digital language infrastructure for Dutch.

In this paper we report on the past 10 years of HLT Agency activities and look to the future. In particular, we summarise how and why it has been set up (section 2) and which are its most important achievements so far (section 3). In view of the next decade, the role of the HLT Agency is reviewed in the following section (4), leading to some operational challenges to cope with and opportunities to take advantage of (section 5). In a subsequent section (6), a more policy and governance oriented discussion is presented, followed in section 7 by some lessons learnt over the years. An innovative outlook to the future (section 8) precedes the conclusions of section 9

2. Sustainability of a local digital language resources infrastructure for Dutch

In many cases, universities or research centres become responsible for the research results generated thanks to government funding at the end of a research project. However, they usually have no funding, official duty or permanent infrastructure to maintain project results and make them available for re-use. Only by attracting new research funding, results of previous projects can be maintained. A lot of uncertainties (funding not granted, researcher leaving the team, low quality of software documentation ...) jeopardise the maintenance and distribution of project results.

To prevent HLTD resources developed with public funding from lying unused on a shelf or on some server in cyberspace, their usability must be safeguarded. This may include debugging or migrating to newer platforms or operating system versions. The NTU, as financer and owner of an important number of HLTD resources, took the initiative to set up the HLT Agency. It has as its mission to manage, maintain and distribute as a "one-stop-shop" digital language resources for Dutch (corpora, tools, lexica etc.) for the benefit of research, education, development and innovation in academia and industry so that Dutch remains a "fully fledged" digital functional language.

Maintaining a digital language infrastructure for a language can be organised at three levels, which are reflected in the operational organisation of the HLT

* As of 2014, the secondment of Peter Spyns to the Taalunie has ended and he is again working full time at the EWI department.

[1] As a consequence of the Belgian state reform (federalisation), Flanders later became the official partner of the treaty.

[2] De Nederlands-Vlaamse Centrale voor Taal- en Spraaktechnologie (TST-Centrale).

Agency. The first level tasks are: acquisition, management, maintenance, distribution, and service delivery regarding HLTD resources. At a next level, the tasks include maintaining and managing the technical infrastructure for executing the first level tasks (including framework conditions such as IPR legislation). And thirdly, it concerns aggregating and maintaining general knowledge on the HLTD resources, on the digital language infrastructure and on the overall HLTD innovation ecosystem (including regularly monitoring the state of the BLARK for Dutch (Daelemans et al., 2005)). Note that he HLT Agency does not produce resources itself nor performs evaluation campaigns. In that sense, it differs from ELDA (Arrantz & Choukri 2010) or LDC (Cieri & Liberman 2010) – see Sect. 4.

It is not only necessary to ensure the sustainability of linguistic resources, but also the technical infrastructure itself must be hosted and maintained on a sustainable basis. It serves no purpose to initiate language resources infrastructure projects if post-project financing is not guaranteed from the start. The NTU, wanting to support and strengthen the position of Dutch in the digital age, funds the HLT Agency on a structural basis for about 450K euros per year so far.

A local resource agency was preferred (to existing international organisations) as it profits from being directly embedded in the HLTD innovation ecosystem in the Low Countries. It is e.g. easier for researchers to interact with the HLT Agency, in particular concerning bug fixes. In addition, the HLT Agency, with its pricing and IPR policies can stimulate – albeit in a modest way – local HLT enterprises. International resource centres are less concerned with the local economy. The HLT Agency can thus become an additional instrument – albeit very specific and of limited scale – for economic development and innovation in SMEs in the Low Countries.

3. Achievements

3.1 Reaching diverse target groups

In the past decade different types of customers contacted the HLT Agency. *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, bible and historical studies, communication and information studies, and Dutch studies from all over the world. Before the HLT Agency existed, researchers often didn't know where to find existing LRs, what they were allowed to do with them and/or had to collect their own LRs before being able to start their research. *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. *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 via the HLT Agency lowers cost barriers for a commercially viable solution. A noteworthy success story is that, as Google has purchased a licence on the Lassy small corpus (van Noord et al. 2013), the HLT Agency became a Certified Google Seller. 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.

3.2 Managing an expanding resource portfolio

Table 1 shows a.o. the increasing number of LRs acquired. The portfolio currently consists of 46% corpora, 28% lexical resources, 14% tools and 12% dictionaries. Also the number of signed licenses is growing. Please note that one signed user license may result in the use of a resource by an entire research group, faculty, university or company. In 2011, a new website with a web shop was created, which had a substantial impact on the number of licenses concluded. The impact of the physical move from the INL to the NTU at the end of 2012[3] and e.g. the fact that INL now host their own service desk for INL-LR related questions are visible as the (temporary) decrease of some indicators. Until 2011 users of web applications had to register themselves. In 2011 we decided to make the applications available without registration. Users simply had to accept an end user license published online. This change of policy was applauded by the public and the use of the applications skyrocketed. Unfortunately, technical problems resulted in the lack of trustworthy user statistics for subsequent years.

Table 1: Some HLT Agency indicators

	2006	2007	2008	2009	2010	2011	2012	2013
# of LRs managed	29	34	44	61	68	72	75	78
Archive Size (in Tbs)	0.1	0.2	0.5	1	1.4	1.5	1.8	2
# signed user licenses	21	98	98	123	109	177	219	172
# for profit licenses	4	4	4	7	6	5	4	9
Gross turnover of for-profit licenses	€ 9250	€ 17350	€ 6600	€ 18350	€ 47270	€ 58550	€ 8250	€ 40450
Registered users of web applications	515	4466	2898	4799	6103	5458	N/A	N/A
Messages in service desk	1420	1565	1408	1629	1648	1972	1420	1349

3.3 Solving IPR issues

During the STEVIN programme, many projects required data and started to negotiate with data providers (publishers, webmasters, students, authors, etc.). To streamline the way intellectual property rights (IPR) were handled, STEVIN's IPR committee and the HLT Agency drew up template agreements for resources acquisition and distribution licenses. STEVIN projects were obliged to use these templates to conclude acquisition agreements with resource providers. Conversely, at the end of a STEVIN project, the consortium handed over the ownership of the

[3] The HLT Agency moved from the INL in Leiden to the NTU in The Hague at the end of 2012. This was a result of changes in the cooperation between INL and NTU.

results to the NTU and signed a user agreement that allowed the consortium to continue to work on the resulting resources. The HLT Agency supported nearly 200 data acquisition processes. As a result, the legal situation and property structure of the STEVIN resources are very uniform and transparent. The HLT Agency hence acquired a very valuable set of resources with an almost unlimited "freedom to operate" regarding the subsequent distribution and exploitation.
The knowledge about (solving) IPR (issues) is reused. Two examples are that the HLT Agency helped a project in building a corpus of written input for children and acquiring lexical data from publishers and another project with drafting a consortium agreement in which the IPR of the partners was properly dealt with.

3.4 Satisfying stakeholders

A simple but strong indication of its success and impact was that in a couple of years after its inception completely from scratch, the HLT Agency had acquired its place in the HLTD ecosystem to such an extent that stakeholders expectations became so diverse and demanding that it looked as if the HLT Agency was already active for many years. Two evaluations of the HLT Agency were organised by the Dutch Language Union, both consisting of a self-evaluation, interviews with selected partners and an online questionnaire. Where the first evaluation (2007/2008) had many suggestions for improvement, the second (2009/2010) showed that the majority of respondents was satisfied and we got 8 out of 10 score from our suppliers.
Also, South-Africa set up a Resource Management Agency (RMA), currently hosted by North-West University's CTexT. The HLT Agency supported the writing of the RMA's blueprint and gave advice on legal and IPR related matters. In 2011, the RMA chose the HLT Agency as their partner for the distribution of South-African language resources outside the continent of Africa. This external token of trust illustrates the quality of the HTL Agency's way of working.

4. Updated role and status in the HLT field

4.1 HLT Agency essentials

Entering its second decade of existence, the HLT Agency has to stay faithful to its original and *fundamental* ***ABC*** while adapting to a fast evolving context:

- Its infrastructure and resources in its portfolio ensure that Dutch remains firmly "***A****nchored*" in the digital world. At regular times, the status of the linguistic resources and services for Dutch is to be checked in the light of new (technology) trends and, if necessary, action is to be undertaken to maintain the position of Dutch as a fully fledged digital functional language.
- Its overall expertise of the materials, of the infrastructure and of the HLT innovation ecosystem in the Low Countries needs to remain a "***B****eacon*" for anyone who needs guidance on HLTD. In one case, the guidance will be limited to forwarding the information seeker to others who are more apt to address the issue; in another case the HLT Agency itself can solve the issue. Knowledge of materials in particular and the HLTD innovation system in general is a requirement for this service.
- Its activities are of a nature that the HLT Agency cannot (and should not) address every issue itself. Hence, it is a necessity to build a "***C****ommunity*" of partners consisting of (commercial) data providers, voluntary source code maintainers, scientific resource donators, paid contractors, knowledgeable civil servants, etc. Stakeholder management is to become an integral part of the HLT Agency activities. Future users or customers of HLTD are to be involved in a process of co-creation when new resources or services are envisaged.

Notwithstanding its particular institutional status (part of an intergovernmental organisation), the HLT Agency has to critically reflect on its *position in the international HLT field* as other repositories for managing LRs exist or emerge (cf. (Mariani et al. 2011, p.39 for an overview). Three major examples are DANS and TLA (in the Netherlands) and ELDA (in Europe):

- Data Archiving and Networked Services (DANS)[4], an institute of the Royal Netherlands Academy of Arts and Sciences (KNAW)[5] and the Netherlands Organisation for Scientific Research (NWO)[6], 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 Language Archive (TLA)[7], a unit of the Max Planck Institute for Psycholinguistics, is concerned with digital language resources and tools. It archives resources on languages worldwide (especially endangered languages) and is involved in many (infrastructure and software development) research projects. The TLA is funded by the Max Planck Gesellschaft[8], the Berlin Brandenburgische Akademie der Wissenschaften[9] and the KNAW.
- The European Language resources Distribution Agency (ELDA)[10] is the operational body of the European Language Resources Association (ELRA)[11] and was set up to identify, classify, collect, validate and produce language resources. ELDA is also involved in HLT evaluation campaigns.

Whereas DANS supports all research areas and TLA and ELDA LRs in many languages, it is the HLT Agency's specific mission to take care of digital *Dutch* LRs and to ensure that LRs are not only managed and made available, but also kept *up-to-date* and usable in order to strengthen the position of Dutch in the digital age. This means the HTL Agency takes care of the management of the entire lifecycle of LRs, including maintenance and support (van Veenendaal et al. 2013).

4.2 Renewing the strategy

New resource centres are emerging all over Europe thanks to European and national research infrastructure funding

[4] http://www.dans.knaw.nl
[5] http://www.knaw.nl
[6] http://www.nwo.nl
[7] http://tla.mpi.nl
[8] http://www.mpg.de
[9] http://www.bbaw.de
[10] http://www.elda.org
[11] http://www.elra.info

(e.g., CLARIN (Váradi et al. 2008), DARIAH (Blanke et al. 2011) and META-Share (Mariani et al. 2011), most of them specialising on one or more human language (technology) subdomains. The goal of the HLT Agency – to strengthen the position of Dutch in the digital age – is still sound, but the HLT Agency's means may have to be renewed due to the changed and changing landscape. After all, the HLT Agency and these other centres are all funded by public money, which is by definition scarce, in particular in these years of budget reductions and (government) spending/cost cutting. Currently, the exercise of *renewing the mission statement*, creating a new vision and societal aims, and defining operational targets is on-going. At the time of writing, the following additional roles, rooted in the three layers of operational tasks (cf. Sect. 2), are being discussed (see Figure 1):

1. "the HLTD Monitor": collect and aggregate specific information on actors in the HLTD field in the Low Countries and monitor the status of the BLARK for Dutch, relevant for policy purposes;
2. "the HLTD Services Agent": provide information and advice on HLTD in the Low Countries in general by the combined means of a newsletter, web pages, a humanly operated service desk, a (nearly) exhaustive on-line overview of HLTD resources and organisations, and personalised advice etc.;
3. "The HLTD Marketeer": promote HLT(D) in general and in particular the resources and services offered by the HLT Agency (which can include resources owned by third parties);
4. "the HLTD Resources Tailor": take care of on demand customising and tailoring of the resources managed by the HLT Agency;
5. "the HLTD Tutor": provide (modern) course material on (the most important) resources in the portfolio of the HLT Agency as well as on standards adhered to by e.g. the CLARIN-network.

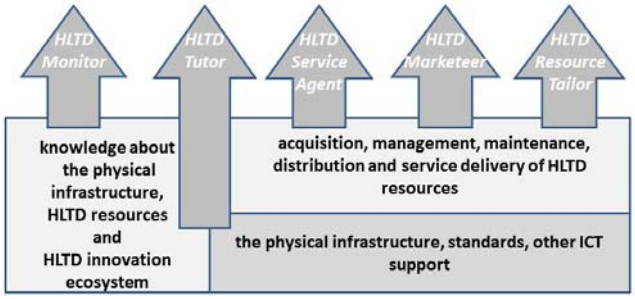

Figure 1: roles and tasks of the HLT Agency

A particular stakeholder is the Dutch Language Union as language policy maker for Dutch and most important funder of the HLT Agency. Both parties have to continue to regularly align their vision and strategies, as the HLT Agency is supposed to operationalise and implement important parts of the overall language policy defined by the Language Union. Both entities are defining their social return on investment and describing convincing societal business cases to substantiate their claim for funding by the Flemish and Dutch governments.

5. Operational challenges & opportunities

5.1 Acquiring new materials

An important challenge for the HLT Agency is to maintain a steady influx of new materials after the end of the STEVIN programme and set up a policy to attract and select new materials. As the HLT Agency operates as a "multi-sided platform" [12], additional, complimentary services to resource providers must be put in place to "reward" providers for their deposit. E.g. in Flanders, a list of high impact journals in Dutch typical of the humanities and social sciences has been compiled for the purpose of citation impact assessment for research funding.[13] At that occasion, it was decided that corpora, collections and similar types of resources with an ISBN number can be cited and hence taken into account for bibliometric impact studies. Researchers thus very much appreciate the HLT Agency becoming a publisher of ISBN numbers and registering an ISBN number for their deposited resources.

One can also think of setting up an upload and download site (a kind of specialised source forge) for researchers to post and download 'as is' demonstrators, resources and source code, whereby a clear position statement is needed w.r.t the HLT Agency's software sustainability policy. This could be linked to a web hosting platform for running demonstrators. E.g., all too often, researchers make nice or interesting demo's that hardly survive the "life time" of a project or a researcher at his/her institution of the moment. Minimally a security check of the demonstrator software must be performed by the HLT Agency, protecting the software (and the HLT Agency's servers) against hackers.

A simple registry with URLs pointing to resources available elsewhere (i.e. not managed and distributed by the HLT Agency) can help generate web traffic to that web site in order for the depositor to gain visibility and fame. Another, simple to apply "rewarding system" is to include a his/her name and/or link in a newsletter, announcement block on the home page, twitter messages, blogs etc. of the HLT Agency and /or NTU. Many more reward systems are possible and the most promising will be selected and implemented by the HLT Agency.

5.2 Widening the scope

In line with the open data idea, funding agencies should point funded researchers to the possibility to deposit their HLT results and resources with the HLT Agency – or even better – allow funding money to be used for maintenance by e-repositories [14]. In Flanders, contrary to the Netherlands, there is no e-repository such as DANS. As some researchers are less forthcoming to deposit their resources at a purely Dutch institute, the HLT Agency, as a country neutral party, can turn this into an opportunity to open up its services from purely HLT to the social sciences in general (albeit only for resources concerning the Dutch language). Conversely, researchers from the Netherlands can positively express their desire to share their resources

[12] "Multi-sided platforms bring together two or more distinct but interdependent groups of customers. Such platforms are of value to one group of customers *only* if the other groups of customers are also present" (Osterwalder et al. 2010, p.77). In the HLT Agency context, it is not difficult to understand that resource users are "connected" via the HLT Agency infrastructure with resource providers (and vice versa).

[13] http://www.ecoom.be/en/services/vabb

[14] This was already mentioned in the HLT Agency blueprint.

with the entire Dutch language area by depositing their materials at the HLT Agency. The latter is supposed to promote the materials in its portfolio in the Netherlands, Flanders and South Africa whereas national repositories only cover the national territory.

5.3. Retaining derivatives

One of the most distinctive features of the STEVIN programme was the transfer to the NTU of the ownership of the STEVIN resources. User licenses include a right of first refusal for the NTU. The latter means that a licensee of a NTU resource at the HLT Agency is obliged to offer any derivatives to the HLT Agency for distribution first.[15] Thanks to this legal arrangement, it is easier for the HLT Agency to keep track of and manage derivatives and updates of the resources in its portfolio.

Upcoming infrastructure centres or networks, in particular those that promote open source licensing schemes, are struggling how to "retain" derivatives and updates of the resources made available within their centre or network, because most open source regimes grant users the right to distribute the resource, derivatives and updates themselves. This freedom is in itself a good thing, but ensuring that the latest version of a resource or derivative is and stays in a certain centre or network greatly simplifies the resource's life cycle management. Many more distribution licenses pose restrictions on redistribution of the resource, derivatives and/or updates (e.g. Oksanen et al. 2010, Choukri et al. 2011)[16].

5.4. Promoting standards

With the advent of linguistic infrastructure networks, an opportunity arises to shape the HLT field in terms of de facto (or somehow agreed upon) standards. Scientific literature on open innovation theory states that standards are an important means to boost research and innovation in a research domain and/or economic sector, hence the HLT Agency can become an influential player in the HLT landscape in the Low Countries and pivot between academia and industry (e.g., by validating submitted resources against the relevant standard(s) – or in a weaker form guidelines – as part of a quality control cycle). It is also expected that the CLARIN network becomes a major driver – e.g., by adopting the Data Seal of Approval. The HLT Agency is also in contact with a Dutch digital humanities software sustainability alliance initiative. Note that on this issue companies in particular prefer resources to comply to (open, industry) standards.

5.5 Building strategic intelligence

Notwithstanding its participation in international networks, the HLT Agency must build up its own strategic intelligence and forward looking capacity rooted in the Flemish and Dutch contexts. External experts are invited for their advice about how to manage and extend the portfolio on resources and services: themes such as pricing policy, partnerships with industry as well as upcoming trends, governmental policies in the making, and expected local needs for resources and services by academia and industry are covered. Of course, various types of local stakeholders should have the opportunity to enter in a dialogue with the HLT Agency to communicate their wishes and expectations. In the long run, the HLT Agency will thus be able to provide more strategic advice to customers and stakeholders, which are not always acquainted well enough with these matters. E.g., SMEs are not always aware of opportunities for (public-private) funding and cooperation.

5.6. Reaching out to industry

Currently, the HLT Agency has some difficulties in reaching out to private enterprises. Most "hard core" Dutch HLT developing companies, in particular those involved in the STEVIN programme and those connected to the Dutch Organisation for Language and Speech Technology (NOTaS) [17], are aware of the resources available at the HLT Agency. In Flanders, the HLT Agency and its materials are not so well known. Extra marketing activities and more intense contacts with what are called "intermediaries" is required. The HLT Agency has begun to connect itself to sector organisations and innovation networks (in many cases supported by the government), ICT incubators and/or entrepreneurial stimulating platforms – e.g., electronic platforms where technology requests and offers or partner searches can be posted.[18]

In addition, a continuous challenge for the HLT Agency is how to provide specific, tailor-made services to private enterprises. In general, the idea is that companies can improve their competitiveness with innovating products and services thanks to LRs. As LRs have been created for general purpose use, the demand of companies is not always matched by what the HLT Agency has in store. On the other hand, in many cases it is not feasible for the HLT Agency to deliver custom made data sets on demand and at short notice. However, the HLT Agency can support companies by offering more interesting licensing conditions. Some companies want to examine more data than the sample made available under an evaluation license before buying the entire set. Others have asked for an – originally not foreseen – research purpose only, cheaper license than the unlimited but more expensive commercial license with a lump-sum price tag. Royalty-based licensing is also an opportunity, e.g. for start-ups who are unable to pay the lump-sum prices on beforehand. It is a healthy strategy to try to achieve a win win situation with (potential) customers by remaining attentive to their requests.

6. Discussion

6.1. Networked R&D&I

At its origin, the HLT Agency was in practice an atomic entity that got all its initial resources (corpora and dictionaries) from its two most related organisations, namely the NTU and the INL. Interested parties from the Low Countries had to download, fill out and return a signed paper form before receiving the data shipped by mail on CD-ROMs. Later on, a web shop allowed

[15] The ownership of the derivatives remains with the developer.

[16] If the original developer agrees to this.

[17] http://www.notas.nl

[18] Examples on the European level are the Enterprise Europe Network: http://een.ec.europa.eu/ and the LT Innovate organisation: http://www.lt-innovate.eu/ .

customers to order a resource by clicking and uploading a signed license agreement in order to swiftly receive a download link to access the resource. In the (near) future, web services could be made available over secure network infrastructure (e.g. by CLARIN) to process 24/7 on the fly requests from all over the world. These web services are in changing configurations part of various larger value (processing) chains (e.g. work flow systems, linguistic on-line work benches, corpus processing pipe lines, machine learning based applications training and customisation environments). From a commercial point of view, one could image to charge companies with a small fee per access (micro payment) to a resource ("resources on demand") or for the use of a resource for a certain time: hiring instead of acquiring. New internet business models are nowadays focusing on paying for permanent availability of and access to resources instead of selling copies of resources.

6.2. Innovation ecosystem

Next to the hardware and software aspects of the linguistic infrastructure, a crucial asset, sometimes overlooked but nevertheless also part of the overall linguistic infrastructure, is the strategic intelligence of the field or systemic expertise. More precisely who is doing what and who has which knowledge and expertise. A role for government (supported) organisations can consist of detecting newcomers and keeping track of existing actors in the HLT domain. According to the theory on innovation systems, more and more focus lies on the links and connections between the various actors of the system. Governments can effectively contribute to the dynamics of the HLT innovation system by organising and/or supporting match making and networking events and activities. Others would call this community building on the one hand and activating the wisdom of the HLT crowd on the other. E.g., policy makers could point actors on upcoming (funding) opportunities or ask for advice when preparing new initiatives. On a lower level, such activities (extending into co-creation activities) can be organised around certain linguistic resources by means of specific implementations of linguistic infrastructure. Which closes the circle or rather creates an upwards spiral. E.g., the HLT agency is discussing an agreement with the not for profit organisation OpenTaal that allows the latter to distribute as open source a synonym list (derived from the SoNaR Corpus – a STEVIN project (Oostdijk et al. 2013)) for usage in LibreOffice. In return, OpenTaal could promised to maintain and extend for free the synonyms list thanks to its many volunteers.

6.3. Market stimulation

However, the HLT Agency should take extreme care not to distort the market and become a government funded competitor of private enterprises. The NTU and/or the HLT Agency can step in when it has become clear that industry is not willing to risk an investment in certain resources or tools (market failure). The results of such an intervention must be made available on equal terms to all (industrial) players. E.g., as nowadays almost any software engineer can develop mobile apps (cf. the growing number of hackatons in all kinds of application areas), it is tempting to allow free and unlimited use of LRs for Dutch as the development of free apps can be seen as a strategy to support and strengthen the position of Dutch in the digital age (language policy goal). On the other hand, such an app can ruin the market position of an existing local company that also receives government money to innovate its products and services (innovation policy goal). It is a delicate exercise on how to reconcile both policy measures that apparently are sound in isolation but nevertheless may conflict in a broader context.

6.4. (Local) language policy vs. (global) science policy

Synergistic use of human resources and other means is efficient and cost-reducing, compared to, for example, having several (smaller) organisations that only take care of certain aspects of the LR lifecycle instead of its complete lifecycle. Furthermore, combining resources and bringing together different kinds of expertise on LRs creates a surplus value. This can result in improved versions of datasets and new insights into potential use(r)s of LRs. Although the HLT Agency is of a much more modest dimension than LDC and ELDA, policy makers in the Low Countries consider it worthwhile to structurally fund a local linguistic resources and services centre that covers the national language. Having one organisation that is responsible for managing LR lifecycles creates higher visibility of and better accessibility to these resources. An important difference with most of the resources centres is that the creation of the HLT Agency is an integral part of a national language policy rather than an international research infrastructure policy. Even if many aspects of governance, practical set up and effects are alike and can lead to synergies, this fundamental difference is reflected in a completely different decision rationale regarding the funding. International infrastructure networks will ultimately be judged (by the participating (inter)national funding organisations) on its capacity to enable cutting edge research (e.g., CLARIN) or to stimulate HLT R&D (e.g., META-Share). At regular intervals, a renewal of the funding will be subject to evaluation and impact assessment in line with the most recent overall policy (e.g., Horizon 2020). As Dutch is not a what is called "top-10 language" long term funding for a digital language infrastructure for Dutch may be more "sure" than renewable funding for research infrastructure.

Not only can the HLT Agency thus be seen as an expression of culture pride and ambition of the Dutch language. In addition, researchers and industry representatives in the Low Countries appreciate the nearness and direct contact that favours knowledge spill over (in both directions). A linguistic resources and services infrastructure centre embedded in the local innovation system is more beneficial to stimulate open innovation and co-creation activities, in particular when a shift towards services is made (see Sect. 8). It is revealing in this respect that within the CLARIN ERIC, the most recently defined centre category is that of the L-type, meaning a knowledge centre that offers resources and services specifically to a *local* community (in contrast to a K-centre that offers specific expertise that is relevant to the entire CLARIN network[19]).

[19] Evidently, this implies that not too many centres can offer the same or overlapping type of expertise.

6.5. Role in the CLARIN network

In section 4, we have already touched upon the possible roles of the HLT Agency in the CLARIN network. By taking up these roles, the HLT Agency helps in putting together a global CLARIN Knowledge Sharing Infrastructure (KSI). As the HLT Agency primarily targets the Low Countries, it evidently, already now functions as an L-centre. In addition, once the web shop, resource documentation, PR-material, course material and the like have been translated to English, the HLT Agency can easily become a CLARIN K-centre.

In line with the overall NTU policy of co-operation with South Africa, the HLT Agency has concluded a collaboration agreement with the South African Resource Management Agency, thus adding an international dimension – even if a lot of translation work still has to be done. An HLT Agency "going international" also supports other NTU policy choices. E.g., students abroad studying Dutch or foreign research groups working on HLT can profit from the HLT Agency. Conversely, HLTD resources created abroad can become part of the BLARK for Dutch (and possibly integrated in the HLT Agency portfolio). HLT roadmap planning, language policy preparation and joint R&D or infrastructure planning on an international level can benefit from knowledge accumulated by the HLT Agency. Hence, it is merely a matter of time and human resource effort to slowly but surely have the HLT Agency evolve towards a CLARIN K-centre. E.g., we don't expect that other CLARIN centres will track the status of the BLARK for Dutch or aggregate information on the HLTD innovation ecosystem. Hence, by contributing to the CLARIN ERIC KSI, the HLT Agency and the NTU also achieve some of their other policy goals.

7. Lessons learnt

In the course of the past decade, the HTL Agency has become an important actor in the HLTD innovation ecosystem. However, it became clear that a *longer term vision* that is *shared* and commonly agreed upon between the NTU (the funder and language policy maker) and implementation party (HLT Agency) is of crucial importance. It is in the interest of all parties to commit to clearly defined roles, tasks and responsibilities and a more in-person cooperation relationship. The HLT Agency's move from the INL to the NTU has proven instrumental in this respect.

It was complicated to draft IPR agreements that were happily accepted by all stakeholders. Many researchers prefer open source as a legal framework – albeit sometimes for the so called reason of avoiding administrative overhead. However, from the point of view of a resources maintaining organisation that should also service industrial partners, acquiring *the ownership of the materials* – as was the case with the STEVIN programme – opens up many more legal possibilities to exploit the materials. Companies usually prefer an identifiable party to negotiate with and a transparent ownership structure of the resources. The ownership of the HLT Agency materials is legally sound, the terms of (re-)use are unambiguous and apply to any subsequent user of the resource without him/her having to worry whether or not parts of the resource have been obtained illegally – as happens in quite often in "web harvesting" projects. This guarantees a legally stable situation for all licensees and providers.

8. Outlook

Funding schemes for setting up language infrastructure centres and networks have become quite popular in recent years (CLARIN – Váradi et al. 2008; META-Share – Mariani et al. 2011). Good descriptions of "the state of the official language(s)" of a country are needed. Such reports are available for quite some languages[20] – e.g., (Odijk 2012) for Dutch. A BLARK – e.g., for Dutch (Daelemans et al., 2005) – is the technological translation of the status of a language. A BLARK in essence is a set of linguistic resources that should be minimally available for a language – e.g., corpora, dictionaries and tools. It also identifies HLTD resources that are still to be developed. As such, a BLARK is an useful instrument for language policy makers.

As a successor to the BLARK, national language institutes might think of the next step, which we coin *BLAISE* (Basic LAnguage Infrastructure SErvices). The BLAISE is a collection of services that a language should offer to its (daily) users, developers, researchers and government. In this paper, we presented several possible services. It is up to the language policy makers to determine, together with relevant societal stakeholders, which language related services are minimally required. Technological services can include web services, e.g. in the context of the CLARIN network. One web service development example is the combination of a terminology extractor for Dutch made available as a web service linked to a terminology management service with cloud storage. It supports local translators and terminologists in their research (scientific purpose) and professional activities (economic purpose) while indirectly stimulating the development of Dutch terminology (language policy purpose). Similarly one can think of a 24/7 modular spell checking web service in the cloud that has been certified as the official reference spell checker for Dutch, or other LR reference quality checking (web) services. Of course, the idea of building linguistic web services is not all that novel. However, determining within the context of a national language policy which are the minimally needed services to support the use of that language is a new exercise and, as far as we know, still to be performed.

The BLAISE should also include humanly operated /delivered services such as continuously monitoring the state of the national language infrastructure, giving advise on how to use the language infrastructure's components in research and development and stimulating interactions between all the players in the field (cf. Figure 1). Currently, language policy makers are largely focused on various aspects of language learning and on providing resources to that aim. Even though these aspects remain important, language policy is not isolated from the overall societal context. A more systemic analysis is needed focussing on the roles, functions, etc. of all relevant actors, how to interconnect and strengthen them. A BLAISE is

[20] See http://www.meta-net.eu/whitepapers/overview

part of the outcome of such a systemic reflection on language policy also taking into account issues related to science, innovation and societal challenges.

9. Conclusions

Notwithstanding the particular case of the HLT Agency, the issues raised in this paper undoubtedly play for other language resource centres as well. By teaming up with partners of the CLARIN-network, the perspectives as well as the challenges to the HLT Agency have broadened considerably. Gradually transforming from a BLARK-based agency for government-funded HLTD LRs into a BLAISE-driven organisation with a portfolio of basic language services for various types of users seems a promising future for the HLT Agency. It could enable the HLT Agency to support and strengthen the position of the Dutch language in the digital age even more. The co-operation agreement concluded with the South African Resource Management Agency proves that the expertise and method of operation of the HLT Agency is appreciated not only by stakeholders in the Low Countries but also at the other side of the globe. Hence, it is safe to state that the HLT Agency continues to fulfil a set of useful and necessary tasks for the status of the Dutch language, which are in some cases, as a consequence of its focus on language policy, unique and this even more in an international context.

10. Acknowledgments

We'd like to thank our colleagues at the NTU and the HTL Agency, as well as the external experts who gave feedback on our plans during the strategic brain storm sessions.

11. References

1. Arranz V. & Choukri K., (2010), *ELRA's services 15 years on … sharing and anticipating the community*. In Proceedings of the *7th International Conference on Language Resources and Evaluation (LREC10)*
2. Blanke T., Bryant M., Hedges M., Aschenbrenner A. & Priddy M., (2011), Preparing DARIAH, in *Proc. of the 7th IEEE International Conference on e-Science, IEEE*, pp. 158-165
3. Beeken J. & van der Kamp P., (2004), The Centre for Dutch Language and Speech Technology (TST Centre), in *Proceedings of the 4th International Conference on Language Resources and Evaluation (LREC04)*, pp. 555-558.
4. Boekestein M., Depoorter G., & van Veenendaal R., (2006), Functioning of the Centre for Dutch Language and Speech Technology, in *Proceedings of the 5th International Conference on Language Resources and Evaluation (LREC06)*, 2303 - 2306
5. Choukri K., Piperidis S., Tsiavos P. and Weitzmann H., (2011), META-SHARE: Licenses, Legal, IPR and Licensing issues, META-NET Del. D6.1.1.
6. Cieri C. & Liberman M., (2010), Adapting to trends in language resource development: a progress report on LDC activities, in *Proceedings of the 7th International Conference on Language Resources and Evaluation (LREC10)*
7. Daelemans W., Binnenpoorte D., de Vriend F., Sturm J., Strik H. & Cucchiarini C., (2005), Establishing priorities in the development of HLT resources: The Dutch-Flemish experience, in Daelemans W., du Plessis T., Snyman C. & Teck L. (eds.), *Multilingualism and electronic language management: Proceedings of the 4th International MIDP Colloquium*, pp. 9-23, Van Schaik
8. D'Halleweyn E., Dewallef E., & Beeken J., (2000), A Platform for Dutch in Human Language Technologies, in *Proceedings of the 2nd International Conference on Language Resources and Evaluation (LREC 2000)*
9. Krauwer S., (2003), The Basic Language Resource Kit (BLARK) as the first milestone for the Language Resources Roadmap, in *Proceedings of the International Workshop Speech and Computer*
10. Mariani J., Choukri K. & Piperidis S., (2011), META-SHARE: Constitution, Business Model, Business Plan. META-NET Deliverable D6.3.1
11. Odijk J., (2012), The Dutch Language in the Digital Age, The META-net White Paper Series, Springer
12. Oostdijk N., Reynaert M., Hoste V & Schuurman I., (2013), The construction of a 500-million-word reference corpus of contemporary written Dutch, in *Spyns and Odijk 2013*, pp. 201 – 226
13. Oksanen V., Lindén K. & Westerlund H., (2010), Laundry symbols and license management: practical considerations for the distribution of LRs based on experiences from CLARIN. In *Proc. of the LREC10 workshop on Language Resources: From Storyboard to Sustainability and LR Lifecycle Management*
14. Osterwalder A., Pigneur Y. & Smith A., (2010), *Business Model Generation*, Wiley
15. Spyns P. & D'Halleweyn E., (2012), Smooth Sailing for STEVIN, in *Proceedings of the 8th International Conference on Language Resources and Evaluation (LREC12)*, pp. 1021-1028
16. Spyns P. & D'Halleweyn E., (2013), Joint Research Coordination and Programming for HLT for Dutch in the Low Countries, *Journal of Linguistic Resources and Evaluation*, 47(2): 565 - 574
17. Spyns P. & Odijk J. (eds.), (2013), *Essential Speech and Language Technology for Dutch: resources, tools and applications*, Springer, [link.springer.com/book/10.1007/978-3-642-30910-6/page/1]
18. Váradi T., Krauwer S., Wittenburg P., Wynne M. & Koskenniemi K., (2008), CLARIN: Common Language Resources and Technology Infrastructure, in *Proceedings of the 6th International Conference on Language Resources and Evaluation (LREC'08)*, pp. 1244 – 1248
19. van Sterkenburg, P. Kruyt, T., Van der Kamp, P. & Binnenpoorte, D. (2002) *Blauwdruk voor onderhoud, beheer en distributie van door de overheid gefinancierde digitale materialen*.[in Dutch]
20. van Noord G., Bouma G., Van Eynde F., de Kok D. & van der Linde J., (2013), Large Scale Syntactic Annotation of Written Dutch: Lassy, in *Spyns and Odijk 2013*, pp. 147-164
21. van Veenendaal R., van Eerten L., Cucchiarini C. & Spyns P., (2013), The Dutch-Flemish HLT Agency: Managing the Lifecycle of STEVIN's Language Resources, in *Spyns and Odijk 2013*, pp.381 – 394