

# BUILDING AN INTEGRATED PAN-EUROPEAN NEWS DISTRIBUTION NETWORK

---

Markus Schranz

*Research & Development, presstext Nachrichtenagentur GmbH, Vienna, Austria  
schranz@presstext.at*

Schahram Dustdar

*Vienna University of Technology, Institute of Information Systems  
Distributed Systems Group, Vienna, Austria  
dustdar@infosys.tuwien.ac.at*

Christian Platzer

*Vienna University of Technology, Institute of Information Systems  
Distributed Systems Group, Vienna, Austria  
cplatzer@infosys.tuwien.ac.at*

*News Agencies, like any other provider of electronic services, start to make heavy use of modern Internet technologies to provide fast, reliable and user-oriented services for their customers. To meet today's requirements it is becomes necessary to not only improve these services but also enhance the quality of the underlying methods to create, modify and distribute news articles all over the world. In this paper we discuss an approach to integrate services from already existing news agencies, and build a platform that enables international news agencies to take advantage of the extended functionality and information. Furthermore, we discuss the Web Service-based peer-to-peer network NEDINE, which will act as a case study for the presented approach.*

## 1. INTRODUCTION

State-of-the-art Internet technologies are nowadays used to reach billions of people all around the world. Especially in the media-domain, these technologies pose a valuable tool to establish e-businesses which are able to keep pace with the emerging need for global information.

Electronic publishing has introduced a wide application range within the information management field, including the digitisation of ancient archives, the manageability and availability of enormous amounts of data or the visualization and provision of contents to the research world and the open public [Paepen, 2002; Schranz, 2004]. Based on research in areas like information retrieval, eLearning, security, privacy and especially the Semantic Web, information management systems take advantage of results and enrich their functionality to create modern solutions [Berners-Lee, 1998].

Beside the large variety of scientific approaches, serious business impacts have been provided by Internet technologies. Distributed computing algorithms, the use of modern Internet protocols, and adequate content types have created complex and powerful web-based services [Kirda, 2001]. Utilizing such results, online

information like news have been spread with a new class of speed and range all over the world, making news available to virtually everyone within a very short time.

Recent research work within the EU-funded project NEDINE[Nedine, 2004] focuses on the creation and the conception of an intelligent news publishing and distribution network consisting of existing local news agencies that use modern Internet and distributed computing technologies to build up a new kind of multinational and multilingual news distribution service. Research fields such as network communication [Yu, 2004; Birman, 2003] have been involved to create a scaleable peer-to-peer architecture, artificial intelligence is utilized to identify the most relevant related articles within the entire multinational network that can add to the quality of the business news currently in focus. The vector space model [Salton, 1983; Wong, 1985] has been utilized to provide easy access to related and most relevant business news articles within a multilingual and multinational context.

This paper is focused on the integration of modern Internet technologies such as Web Services [Haas, 2002] and peer to peer architectures [Birman, 2003] to create a scalable and high traffic information exchange and distribution network. A brief introduction to the investigated domain is presented in section 2. Requirements analysis and the systems architecture developed within the research work of this project is described in section 3. As proof of concept the prototype implementation and innovative features are presented in section 4. A short summary and an outlook to further developments, practical and commercial use concludes the article.

## **2. INTEGRATING EXISTING SERVICES TO A PAN-EUROPEAN NETWORK**

Throughout an entire decade, modern Internet technologies have brought unmanageable amounts of information to the average user's fingertips. The information business has gone through dramatic changes in the last century and with the potential range of more than 800 million users [Internet World Stats, 2005] the management of mass information dissemination has become a challenging social and economical task.

The application domain we are focusing on has built a business on top of the technological features and basic services available on the Internet: Several PR-companies and news agencies in Europe have utilized modern synchronous and asynchronous Internet technologies such as web-based information access frontends for business news archives or email-based mass news distribution or content provision for next generation user devices. A media industry has emerged that uses complex and technologically challenging Internet services to create, aggregate, exchange, publish and distribute current business news.

European business today is highly segmented and widely unrecognised beyond national borders. Business news mostly bear national relevance but hold the potential to spread cooperation opportunities and business changes towards an economically and socially integrated Europe.

The discussed project is initiated to build a network to integrate multiple European national information sources consisting of participating news agencies, PR agencies and independent journalists into an international information service for news professionals and decision makers. Existing services are based on modern Internet technologies like Web application servers holding the business logic in the

services middleware [Cvitkovich, 2005] and maintaining a local network of content providing editors and commercial customers as well as thousands of subscribers and readers. Each business news service provides a set of features to these specific user groups (see table 1) that need to be normalized and shared within the integrated network of media partners.

Table 1. User groups and services within the application domain.

<b>Users</b>	<b>Services</b>	<b>Interface Technology</b>	<b>Scale</b>
Journalists and opinion leaders	Free business news	Access via modern web archives, XML-based mobile devices and standard electronic mail	Hundreds of thousands
Editors	Business news creation and aggregation	Modern web forms and interactive web content management systems	hundreds
Distribution customers	Business news dissemination	Customer-oriented web content management interfaces	thousands
News agencies	Business news exchange, aggregation and dissemination	XML-based application interfaces, Web Services	dozens

In several European countries Internet-based businesses following this model have emerged recently. Technologically, the implemented services lack homogenous implementation models, data structures and communication protocols. With modern Internet technologies like Web Services for the information exchange and peer to peer architectures to manage a scalable integration of several local service providers a meta-network has been designed within the scope of the project NEDINE.

The project aims at providing a network of news exchange and distribution that supports mutual awareness of relevant topics and information areas within all European countries. With its main focus on widespread availability and affordability for all partners it addresses news providers to transport national and international information to the relevant target group, regardless of the origin, nationality and financial capability of the information provider.

The network discussed within this paper is targeting at a multinational and multilingual integration of such business cases, thus allowing the news agencies of different European countries to share their contents and exchange their business news towards an integrated network for news aggregation, creation and dissemination. Aside from the obvious business benefits of such a service integration, there are necessary steps to be taken to technically and organizationally bring the services and the systems of the existing news agencies to an integrated network. Beyond this, the project consortium is currently developing a demonstrator and initial business service that shall attract additional agency partners throughout Europe to join the network in order to have access to relevant business news at an international level and to offer a distribution and dissemination interface for their customers that provide news to the network.

### **3. SYSTEM ANALYSIS AND DESIGN**

Internet technologies in the area of business news distribution involves technical features to manage scalability and performance in mass information provision (millions of page impressions per month) and mass distribution (millions of electronic mails sent per day). Since scalability is managed by integrating local strength into a powerful network, modern networking features and capabilities are within the focus of the technical architecture.

The goal is to set up a sophisticated news platform and a high performing distribution network based on convenient digital news exchange technologies, not only targeted at newsrooms but aimed at reaching as many individual business leaders and decision makers as possible directly and personalized at their desktop. Experiences from local providers give figures of hundreds of thousands of electronic mails per day and about the same amount of visits to the online business news archives (Web presentations).

Besides the challenging distribution scale, the network shall also integrate independent journalism work and company-oriented business news to form a comprehensive international information service for business leaders. Basic online content management software shall be extended with features for multilinguality, XML-conforming standard formats and personalized distribution. Current distribution technologies have been extended to reach information professionals at multiple end user devices in most current formats. NEDINE has been preparing for different submission channels by using standardized data formats like XML news applications (NewsML[IPTC, 2003], NITF[IPTC, 2000]).

#### **3.1. Requirements Analysis**

In order to build a pan-European network of business news brokers, we have been concentrating on common features and identified technological concepts shared by the participating news agencies within the NEDINE project consortium. The agencies have developed a strong national news network and have been recently developing digital electronic platforms to manage the news provided by own staff or integrated from remote resources. The subscribed content customers utilize their business connection to the agencies to be able to retrieve contents from the digital platforms.

The news agencies manage their content platforms by providing news of different areas of interest via typical Internet protocols and interfaces. All partners utilize Web servers as part of their service, so the integration could be started from this point. External resources like national and international content providers are used by the agencies to enrich the contents for the national platforms.

The integration work was started in the requirements analysis by identifying and describing the existing platforms and services in order to come up with a feasibly information exchange interface and an integrated network service. As a result we have identified the existing services worth being maintained and integrated into an international network of strong local partners.

The integrated news distribution network includes several services that originate from different sources: some are provided by the existing local services and others are provided by the integrated network capabilities and features. An important precondition to facilitate both technical feasibility and business initialisation was to not reinvent existing solutions. In order to fulfil the multilingual and multinational

requirements of the integrated news distribution network, existing solutions needed to be adapted and new software had to be designed and implemented. In order to keep costs manageable and the network attractive for future partners to join, the consortium followed the approach to request as few adaptations to existing services as possible and as much adaptations as necessary.

The requirements focused at the features of the existing services and the necessities of the integrated multinational network. Table 2 briefly shows an overview on the tasks already managed in existing systems and those that have been designed for the integrated network.

Table 2. Requirements analysis – Services locally and in the integrated network

<b>Local system services</b>	<b>Data Design and Processing</b>
News creation and management	Handled in relational database schemata, manipulated by web-based content management engines
News aggregation	Aggregation of third party information (b2b news exchange); Upload of business news by customers; Handled by local protocols, web-based applications
News distribution	Handled by email services to individually managed profiles of readers and subscribers
News presentation	XML-based, mainly web-oriented interfaces to provide free access to business news via Web browsers
<b>Integrated network services</b>	
Multinational distribution	Aggregate business news from other partners and use local distribution mechanisms
International enrichment	Provide most relevant semantically related business news for each article; focus on local strength and specific enrichments per partner
Multilingual news distribution	Provide business news for all subscribers and opinion leaders in targeted business languages
Multinational and multilingual news service	Offering multilingual and international business news tailored to the individual needs of the subscriber for asynchronous or synchronous media access

Technologically, the legacy applications at the existing systems provide only a subset of the required functionality of the designed multinational and multilingual news distribution service and therefore need to be extended with the following unified features: a unified access to local articles, the standardized distribution of foreign articles and a unified presentation of network articles.

### **3.1.1 Unified Access to own articles**

The integration design requires a unified interface to the existing system in order to access a single business news article dedicated to dissemination via the NEDINE network. The defined requirements include (1) the definition/existence of a unique ID of such articles, (2) a URL to retrieve an entire article for NEDINE activities (enrichment, distribution), (3) a unique standardized format for the article content and meta information (XML NewsML resp. NITF).

### ***3.1.2 Standardized distribution of foreign articles***

Each integrated network partner is obliged to offer a distribution service to send all information that is sent to the partner on a distribution request by NEDINE to all interested local subscribers and present the information on the local archive (website or other electronic publishing services). The defined requirements include (1) the interface partners are obliged to accept electronic data from other network partners in order to integrate it with his local distribution service. The interface is based on SOAP and defined in WSDL[W3C, 2001], (2) the data format accepted by existing services is defined to be NewsML resp. NITF; (3) a common branding of NEDINE articles is designed

### ***3.1.3 Unified presentation of network articles***

Multinational and multilingual aspects of the integrated network require additional features from local services in presenting contents to their subscribers and readers: (1) multinational semantic enrichment: the local partner needs to present the automatically created links to related news on its presentation platform by adequate visualisation; (2) local articles need to be provided at foreign news archives. Whenever a subscribers requests one of the related news from the above mentioned enrichment, the article needs to be provided at the site of the requestor, independent of the site of the originator; (3) the interface partners are obliged to accept electronic data from in order to integrate it with his local distribution service is based on SOAP and defined in WSDL.

### ***3.1.4 Specific multinational and multilingual network features***

The integration of multiple local services to a news distribution network involves administrative features that create a integrated service that offers more than just the sum of the individual partner's products. To integrate common services, the system has been designed to be able to (1) register all potentially reached subscribers (in terms of language, topics of interest and quantities) within the local service to a logically centralized component of the NEDINE network; (2) present the range of all reachable subscribers with a single article submission, according to the customer-selected topic areas and distribution languages; (3) offer Multilanguage translation of relevant business news through the local competences of the participating network partners in order to optimally serve the customer.

## **3.2 Architectural details**

The requirements analysis has shown the need for a scaleable integration of exiting systems with little intervention and adaptation to the heterogeneous software components. The final decision taken in the project is based on a peer to peer network which is a unique way to create a network for news exchange between European news agencies. Properly developed state machines and interaction models define the entry point for the implementation that was started to create a first prototype. New technologies like NewsML, Web Services via SOAP or XML in general, have been used to create this decentralized system and connect it to all participants.

With the implementation of the methods described briefly in the requirements analysis, we have established a reliable network which is easy to use and easy to

integrate into the target systems. The core architectural concepts handle data management, the publishing interface (main focus on web technologies using modern web application servers) and the interface connections. The interfaces between the communicating peers and the interfaces between a peer and an existing local system are defined in detail by Web Service descriptions. The WSDL details are out of scope of this paper. Every news agency only sees the SOAP interface of the peer that is assigned to it as shown in Figure 1. Therefore, NEDINE functions as a black box environment for them.

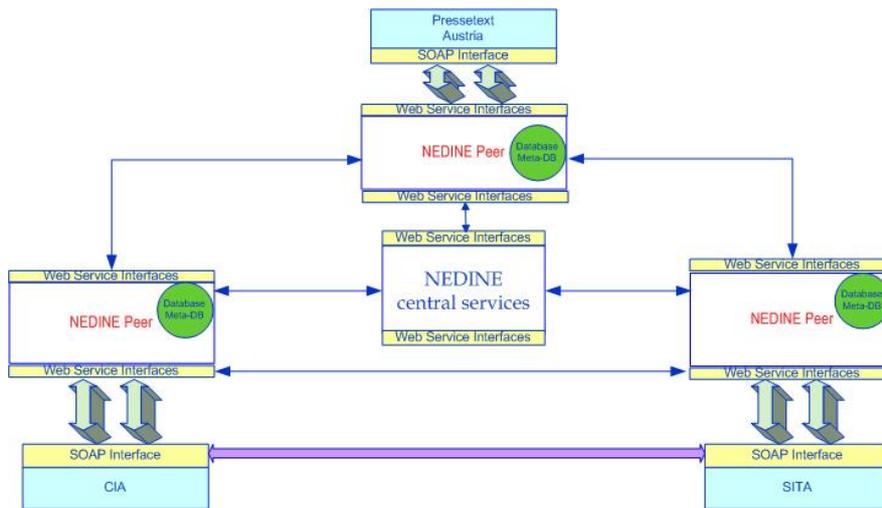


Figure 1: NEDINE P2P architecture – existing agencies and peer components

Existing services use modern web services to retrieve remote business news information to be published by Internet technologies for the local service and/or distributed business news via local channels and the network for remote distribution. Figure 1 shows the existing service as a circle and the newly introduced peer component as an attachment.

The peer to peer concept was chosen since technically all current and future network partners can be equipped with the same software. The network communication can be handled within the peer logic, only the interface between the peer and the existing service needs to be implemented by the joining partner.

All activities within the multinational and multilingual network are triggered by the actions of one of the participating partners, which are represented in the architecture as an existing agency. The communication types within the system are structured in service registration, information upload, news distribution, and news enrichment calls.

### 3.3 Use Cases

Based on the requirements, the technical architecture and the communication types the created NEDINE network architecture can be explained in different use cases and possible scenarios during service utilization. To demonstrate Web services and

peer to peer communication in action we chose a specific scenario on delivering a business news item multinationally in multiple languages.

The originating network partner (in Figure 2 named CIA) sends several messages to its peer, one per language (Czech and English, in this specific case), each with a request for enrichment containing an enrichment flag with a non-existing country and the article in the corresponding language.

The answer to this message will contain a set of articles, those related to the one in the message both locally and in the countries supporting the language of the article, in the same language. That means, after several enrichment processes for each language and each country, the result dataset has to be merged to one final result set.

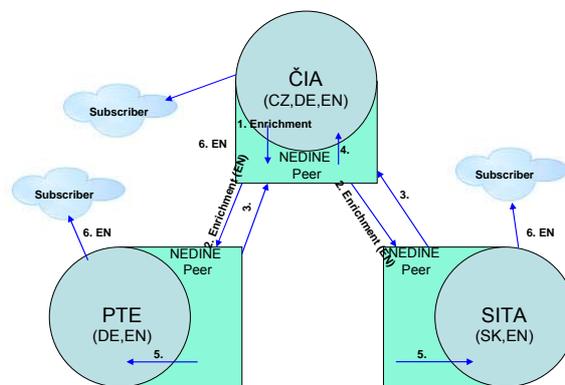


Figure 2: Scenario multilingual and multinational news distribution

The originating network partner sends a message to its peer with a push request containing a push set with all the articles in the different languages and a push flag with the same contents as before (Lang). The answer to this message indicates whether the article has been effectively pushed into the countries supporting at least one language of the article (but in all the languages supported that match with the languages of the article).

The receiving network partners (in the graphic PTE and SITA) send a message to their peer with a polling request and receive pairs of an article (the article in the initial push request, in the use case) and a set of articles, those related to the other article in the pair either locally or in the originating country (in the use case, CZ), in one language. Note that intermixing related articles of different languages is not desired by the news agencies, because of usability reasons.

#### 4. SEMANTIC ENRICHMENT OF BUSINESS CONTENTS

Business news reach a specific audience: researchers that observe markets for specific branches or to analyze modern trends as well as competitors and commercial organizations observing tendencies to optimise their business. Publishers hold the power to bias the market by filtering or targeting the business news towards their own needs. With sophisticated methods of the area of artificial intelligence, automatic content relation management has been applied to the discussed business news publishing and distribution service in order to provide a fair

and informative enhancement to existing solutions. The most relevant semantically related business news applied to the current news item in focus are retrieved and provided via the network service.

For this purpose the practical prototype as a proof of concept for the technical architecture and the semantic coupling of contents uses a search mechanism common in modern information retrieval systems: the Vector Space Model [Wong, 1985]. This approach is mainly used for search engines, based on natural language. Google for example utilizes this method to search their repository of Web pages. The underlying concept is quite simple. Any document is split up in keywords. Each of these keywords constitutes a dimension in a n-dimensional vector space. Therefore a document can be seen as a vector within this "term space". The position of this vector to other vectors within the same vector space describes their similarity to each other. The mathematical method to evaluate how similar two documents are to each other varies. A simple method is to calculate a cosine value for them and express the result as a percentage rating. This method produces very good results for natural language but it is not limited to this field alone.

Virtually any document collection can be mapped to a vector space to create an efficient search environment. In the case of business news publishing and distribution, specific weighting can be applied according to the data formats used within the information network. It is possible to apply phonetic algorithms to methods name. As a result, methods with an equal "sound" produce a high relevance rating. Furthermore, keywords like the business news category or the document title are used to categorize news items according to the assumed behaviour.

Since the Vector Space Model creates high ranking results whenever more terms in the observed article set match, the model provides a neutral and automatically created semantic neighbourhood to the article in focus. Especially in business the most relevant related articles are of high value to the subscribers, since they often want to have a comprehensive overview of a specific market niche or a business area identified or addressed by a single article.

## **5. SUMMARY AND FUTURE WORK**

What we presented in this paper, is a business-oriented network to publish and distribute news over a peer-to-peer network. The Web Service-based architecture guarantees a high interoperability-level with minimum of necessary changes to legacy systems and on the other hand a future-proof and easy method to expand the network to numerous participants when needed. Internet technologies and AI-methods for information retrieval have been investigated and the most feasible approaches have been shown to prove the power of modern research results and to use them to build a successful multilingual and multinational business news platform in Europe.

Apart from the technological challenge, the industrial and commercial sector is an ample motivator, as the described project is funded by the European Commission as an eContent project with strong market orientation. One of the goals is, to apply modern technologies in successful business environments.

Supported by the research experience from project partners, the creation of the final service architecture and interaction model was possible. The result is more than just a junction of networks and services. It is an aggregation of independent

journalists who can present national relevant articles and at the same time, internationally operating companies can use the network to present their information in both, national and international context. The AI technologies used to create the connections between the published articles enrich the content and therefore provide an additional feature for every single participant.

For the future, the existing prototypes will be integrated to a stable network to create an affordable solution for all interested partners in all participating countries. On the business side, the NEDINE project will create a valuable service for news distribution which opens up new markets of existing news providers.

## ACKNOWLEDGEMENTS

This work was partially funded by the EU eContent project NEDINE (News Distribution Network, EDC-22225).

## REFERENCES

1. Birman P., K. IEEE Internet Computing: Peer To Peer - The League of SuperNets. IEEE Distributed Systems Online 4(10): (2003).
2. Berners-Lee, T. Semantic Web Road map, 1998, <http://www.w3.org/DesignIssues/Semantic.html>
3. Cvitkovich, A. & Schranz, M. (2005) Applied Middleware: Object-Oriented Content Management Components with Mason, IADIS International Conference on Applied Computing 2005, Algarve, Portugal, Feb 22-25, 2005.
4. Haas H. World Wide Web Consortium - Web Services, <http://www.w3.org/2002/ws/>.
5. Internet World Stats – Usage and Population Statistics (2005), <http://www.internetworldstats.com/>
6. Kirda, E. et.al. Experiences in Engineering Flexible Web Services, in IEEE Multimedia, V8(1), January 2001.
7. NEDINE – Intelligent News Distribution Network for multinational business news exchange and dissemination, EC eContent project EDC 22225, <http://www.nedine.org/>
8. NewsML, IPTC-NewsML Standard, <http://www.newsml.org>, 2003.
9. NITF, IPTC News Industry Text Format Standard, <http://www.nitf.org/> or <http://www.xmlnews.org/NITF>, 2000.
10. Paepen, B. et.al. OmniPaper: Modern Approaches for an Intelligent European News Archive, Proceedings of the IADIS WWW/Internet2002 Conference, Lisbon, Portugal, November 13-15, 2002.
11. Schranz, M. & Paepen, B. Architecture design and application for an Intelligent Distributed News Archive, Proceedings of the ICC3 8th International Conference on Electronic Publishing – ELPUB2004, Brasilia, Brazil, June 23-26, 2004.
12. Salton, G. Introduction to Modern Information Retrieval, volume 1. McGraw-Hill, Inc., 1983.
13. Wong, W.Z. & Wong, P. and Patrick Wong. Generalized vector space model in information retrieval. ACM, 1985.
14. World Wide Web Consortium (W3C). Web Services Description Language (WSDL) 1.1, <http://www.w3.org/TR/wsdl/>, 2001.
15. Yu, S., Liu, J. & Le, J. Decentralized Web Service Organization Combining Semantic Web and Peer to Peer Computing. ECOWS 2004: 116-127.