Tools for trade analysis and open source information monitoring for nonproliferation

Cojazzi G.G.M., Versino C., Wolfart E., Renda G., Janssens W.

European Commission,
Joint Research Centre (JRC), Institute for Transuranium Elements (ITU),
Via Enrico Fermi 2749, 21027 Ispra VA
Italy

Abstract. The IAEA state level concept envisions an objective based and information driven safeguards framework utilizing all relevant information to improve the effectiveness and efficiency of safeguards. To this goal the IAEA makes also use of open source information, here broadly defined as any information that is neither classified nor proprietary. It includes, but is not limited to: media sources, government and non-governmental reports and analyses, commercial data, and scientific/technical literature, including trade data.

Within the framework of the EC support programme to IAEA Safeguards, JRC has surveyed and catalogued open sources on import-export customs trade data and developed tools for supporting the use of related databases in the Department of Safeguards. The JRC software The Big Table, (TBT), supports i.a.: a) the search through a collection of reference documents relevant to trade analysis (legal/regulatory documents, technical handbooks); b) the selection of items of interests to specific verifications and c) the mapping of these items to customs commodities searchable in trade databases.

In the field of open source monitoring, JRC is developing and operating a "Nuclear Security Media Monitor" (NSMM), which is a web-based multilingual news aggregation system that automatically collects news articles from pre-defined web sites. NSMM is a domain specific version of the general JRC-Europe Media Monitor (EMM). NSMM has been established within the EC support programme with the aim i.a. to streamline IAEA's process of open source information monitoring for safeguards.

In the first part, the paper will recall the sources of import/export trade relevant for non-proliferation and will then illustrate the main features of TBT, recently coupled with the IAEA Physical Model, and new visualization techniques applied to trade data. In the second part it will present the main aspects of the NSMM also by illustrating some of uses done at JRC.

1. Introduction

As reported in [1], the IAEA "…collects and processes safeguards relevant information about a State from a wide range of sources: information provided by the State itself (e.g. declarations and reports); safeguards activities conducted by the Agency in the field and at Headquarters (e.g. inspections, design information verification, material balance evaluations); and other relevant information (e.g. from open sources and third parties)". Open source is here defined according to [2] as "*publicly available information that anyone can lawfully obtain by request, purchase, or observation*". It includes, but it is not limited to media sources, government and non-governmental reports and analyses, commercial data, and scientific/technical literature [1], including trade data.

JRC is investigating information analysis for supporting nuclear nonproliferation since 2007 and i.a. supports the IAEA through the development of tools and approaches in the context of the EC support programme to the Agency. The paper presents two main research areas in this domain and is structured in two parts. In the first part, the paper will recall the trade data sources relevant for non proliferation and will then illustrate the main features of "The Big Table" (TBT) [3] JRC software tool, recently coupled with the IAEA Physical Model [4], and new visualization techniques applied to trade data. In the second part it will present the main aspects of the Nuclear Security Media Monitor (NSMM) [1] also by illustrating some of uses done at JRC.

2. Trade analysis: data, methodology and tools

2.1. Data sources

Data services on global trade referred to in this paper are open source, as described in a data catalogue [5] originally compiled by JRC for the IAEA. The data are available either for free or by a subscription fee. The information provided by trade data services has a regulatory origin as it stems from declarations made by traders to national customs authorities. Customs data are collected at the national level and, by decision of individual States, published in *transactional* or *statistical* formats.

Transactional data are close to declarations made by importers/exporters to customs. Declared data fields subject to disclosure may include:

- A code classifying the commodity traded; e.g., according to the *Harmonized System* (HS) [6] product nomenclature developed by the World Customs Organization.
- Free text description of the commodity.
- Quantity, expressed in weight or number of items.
- Value.
- Date of shipment.
- Country/port of import/export.
- Party names.

Statistical data on trade are derived by aggregating transactions data by country, trade flow (import or export), reference period of time (months, years) and product categories as specified by the HS. As an example, COMTRADE [7], by the United Nations Organisation, provides annual series of trade data for 150 reporting countries. COMEXT [8], by the Statistical Office of the European Union (EU), is a second example. Focused on European reporting countries, COMEXT provides monthly records on EU trade.

2.2. Analysis methodology

Global trade data can be used in the conduct of ***trade case studies***. A case study may start with a piece of information or a hypothesis about trade activities on items subject to safeguards, export controls, or other proliferation-sensitive items. The goal of a case study is to consult web data sources on global trade to:

- Confirm or deny the information/hypothesis;
- Find new, related information.

A preliminary step is to identify clearly items of interest to the case study at hand. This is supported by expertise on items' and *reference documents* where materials and equipment subject to controls are listed, defined and described.

With items of interest identified, it is necessary to *map these items to HS descriptors* in order to retrieve relevant trade data[1]. Mapping items of interest to nuclear-related trade to HS is not trivial, because:

- HS codes only approximate targeted items.
- Exporters may (intentionally or unintentionally) declare trade to customs authorities in HS categories aside from the correct ones.

Trade analysts must determine on a case-by-case basis which HS codes to use to best address the case study at hand. A starting point for the selection of HS codes is given by existing *correspondence tables*. Developed by experts of the HS, these tables map items listed for export controls to HS codes. One such a table is the EU 'Correlation Table' [9] CT) developed and maintained by the European Commission, DG TAXUD: CT maps to HS items listed for export controls [10] by EU Member States.

[1] By definition, trade data referred to in this article do not generally include trade undeclared to customs. Some countries include such trade when discovered.

With HS codes selected, a plan of queries on trade data services is designed taking into account the questions addressed by the case study and the range of services available on trade data. For example, querying for statistical data requires the specification of the following dimensions:

— A reporting country;
— HS codes related to the items of interest;
— A trade flow (import, export);
— A time period.

Queries return the list of partner countries for which trade on the query dimensions exist, specifying value and quantity of the trade over time. Results of queries are presented in tabular form for analysis and interpretation. The criteria adopted for the analysis of trade data are, in general, specific to the commodities at hand. One informative generic criterion is the 'trade unit value', defined as the ratio between value and quantity traded for a given HS, in a temporal point (e.g. month or year), between traders (countries or companies). Data points compatible with the market price of targeted items are highlighted as points of interest. These may confirm alleged information or provide new insights.

2.3. Tools

In this context a software tool has been developed by JRC in support to trade analysis. The tool, called **The Big Table** (TBT) [3], allows analysts to search control lists, identify items of interest to trade-related case studies, and link these to technical documentation and descriptors needed to retrieve global trade data. More specifically, TBT is designed to:

— Browse and search a collection of reference documents listing and describing items relevant to IAEA safeguards, export controls and case studies on nuclear-related trade. Reference documents include: regulatory documents, technical handbooks and the Harmonized System.
— By the above process, select items of interest to specific case studies related to nuclear trade.
— Map selected items to Harmonized System codes. HS codes are instrumental to the retrieval of trade data records.
— Export HS codes in formats suitable to query web trade data services.

In short, TBT offers functionalities for analysts to perform steps preparatory to the retrieval and analysis of data records pertinent to case studies on nuclear trade. TBT does not provide access to trade data services, nor tools to analyze data records. Data retrieval and analysis are performed outside TBT.

TBT, originally developed for IAEA, can serve a variety of tasks and communities underpinning export controls, including the rating of items by licensing authorities and supporting commodity identification for customs controls. It is made available to a number of EU Member States licensing and customs authorities.

An extended version [4] of TBT has been recently developed for the exclusive use by the IAEA to include the IAEA's Physical Model (PM) [11] on the nuclear fuel cycle. Eleven Volumes of the PM have been added to the TBT database, linking the PM to other documents in TBT referring to the same or related items, including nuclear and non-nuclear materials, especially designed equipment and dual-use equipment. Figure 1 illustrates the schema of correspondences established between a generic PM Volume and other documents in TBT. Links to the Nuclear Suppliers Group (NSG) lists (INFCIRC 254 1 [12], INFCIRC 254 2 [13]) were derived directly from the PM text as explicit references exist there. All other links were established on the basis of already existing correspondences in TBT. Through these existing correspondences the PM was linked to Annex II of the Additional Protocol and its handbook (INFCIRC 540c II [14], INFCIRC 540c II HB [15]), the handbooks on the NSG lists (INFCIRC 254 part 1 HB [16], INFCIRC 254 part 2 HB [17]), the Goods Review List and its handbook (GRL [18], GRL HB [19]), the EC Regulation for export controls (EC REG 428/2009 [10]) and the Harmonized System [6]. Also, the PM table of indicators was directly linked to the PM database records in TBT to which the indicators refer.

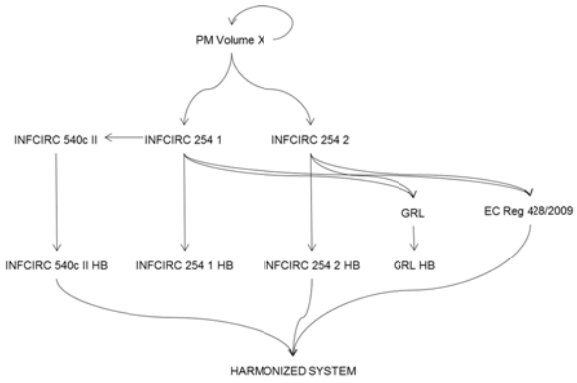

Figure 1. Schema of correspondences between items listed in the Physical Model and items in other documents in The Big Table collection [4].

An issue not completely addressed concerns the linking of manufacturing activities specified in Annex I of the Additional Protocol with manufactured items listed in Annex II to the AP, and NSG 2 (dual-use) items associated with Annex I activities. Preliminary work in this direction is described in [20]. A natural extension to this work would be to create explicit connections between Annex I activities and the PM Volumes. A valuable outcome of the integration of the PM in TBT is the link between the PM and the Harmonized System. This link identifies trade categories that can be used by IAEA analysts consulting global trade data in relation to the PM conceptual framework.

The increasing number of trade categories of potential interest to safeguards highlighted the need to consult and explore the global trade data in a more flexible way than outlined in Section 2.1. This need, coupled with the recent availability of data sets that can be downloaded in bulk, suggested moving to a data warehouse approach where all potentially relevant data are stored in a local database, enriched with meta-data (i.e., product nomenclatures, geo nomenclatures), and then browsed, searched and filtered using data composition and visualization techniques. To this goal, JRC identified the Tableau software [21] as a useful platform to perform exploratory data analysis. JRC is providing support to the IAEA in testing this approach.

2.4 Testing uses of global trade data

Tests on the use of global trade data by the IAEA suggested safeguards relevance along the following lines [22]:

— Support the IAEA State evaluation process and improve the understanding of a State's nuclear programme – Trade information on exports can support the assessment of a State's nuclear related industrial capabilities. Data on trade flows between States can be used to understand their international cooperation. Understanding mining-related activities can be improved by using data on the exports of raw materials and semi-finished products. Data on imports and exports of nuclear materials and equipment may also provide information on the development of the nuclear fuel cycle in general.
— Verify import and export declarations made by States under Additional Protocols (APs), article 2.a.(ix) [14] – Trade data can prove useful to identify flows of raw material subject to safeguards. Trade categories (of the Harmonized System) appear to be less specific than safeguards categories, but precise enough to be determined as safeguards-relevant. The identification of shipments of some AP Annex II equipment may represent a greater analytical challenge.
— Identifying indicators of activities to be safeguarded or to be declared under APs, article 2.a.(iv) [14] – In this context it is foreseen that trade data can be used to verify hypotheses about the absence of undeclared activities. Commodities to serve as indicators and methodologies then need to be identified on a case by case basis and in a hypothesis-specific way.

Actual uses of global trade data at the IAEA are referred in [23].

3. Open source information monitoring: data sources and tools

3.1 Europe Media Monitor Platform and Tools.

As summarised in [1], "the Europe Media Monitor (EMM), developed by JRC [24], is a web-based multilingual news aggregation system that collects over 200000 news articles per day in about 50 languages from more than 4000 web news sources. The sources are mainly general news sites with a world-wide coverage, but also include some specialist websites and twenty commercial news providers. The system employs text mining techniques to provide a picture of the present situation in the world as conveyed in the media. These techniques include automatic multilingual categorization, entity extraction, geo-location, quote extraction and sentiment analysis. In addition, an algorithm for detecting breaking news automatically clusters all collected news articles every ten minutes and displays the ten largest clusters per language by plotting them on a time-by-size graph. Furthermore it applies some deeper semantic information analysis techniques, for example, to automatically detect violent events, derive reported social networks and analyze media impact [25]. EMM creates a searchable full-text index of all articles that flow through the system. For each article, it stores meta-information including title, description, source, category, language, and original URL (Uniform Resource Locator). However, it does not store the original article itself ... The main EMM installation monitors generic news media with little coverage of specialised thematic areas and serves as a general media monitor. Its front page – the EMM Newsbrief [26] - provides a user interface to all this information and is visited on a regular basis by some 25 000 users, and gets some 1.5 million hits per day.". As a complementary tool to EMM, JRC developed the NewsDesk application, which allows an Open Source analyst to manual review and select the most relevant articles collected by EMM. It supports the rapid production of newsletters on dedicated topics which can be disseminated to interested user groups.

3.2 Nuclear Security Media Monitor (NSMM): Extending Source Coverage.

The generic version of EMM monitors pre-selected websites targeting mainly general news media with little coverage of specialised thematic areas. The Nuclear Security Media Monitor (NSMM) is a separate EMM installation dedicated to the nuclear safeguards and security domain [27]. In addition to the generic media sites, it monitors a set of nuclear safeguards-related websites, which were identified and added in collaboration with IAEA in order to increase EMM's impact for the nuclear safeguards analyst. The nuclear sites include: nuclear-focused news agencies and aggregators; regional, national and local government and intergovernmental organisations whose domain covers nuclear issues; NGOs, academic sources and blogs providing analyses on safeguards-relevant topics; general news sources and aggregators; technical publications on the nuclear fuel cycle, etc. A classification of the sources monitored by NSMM is given in Table I.

Table I. Classification of sources monitored by NSMM.

Category Name	Content	Frequency
General News and Aggregators	Directly-accessed news sources, news aggregators, and "fee-based" comprehensive news archive with collection of newspapers, periodicals, and news wires.	Very High
Nuclear News Aggregators	Articles from news agencies and news aggregators that customarily or primarily report on issues related to nuclear industry and safeguards.	High
Blogs	Interactive websites with commentaries on nuclear issues.	Medium
NGO & Academic	Non-governmental organization or university reporting providing detailed reports and added value assessments concerning State's nuclear programmes and activities, and general nuclear nonproliferation issues.	Medium
Government & Intergovernmental	Information from relevant intergovernmental organizations and competent authorities at national level are a unique source of authoritative information on nuclear safeguards and nuclear industry issues.	Low / Medium
Nuclear Industry	Information on companies including location(s), products, capabilities, activities, number of employees, main customers, exports of nuclear related items.	Low / Medium

3.3 Using the Nuclear Security Media Monitor

The two main use cases for the **Nuclear Security Media Monitor** are:

— Real-time Information Awareness. The end user (e.g. an information analyst or nuclear inspector) can directly consult the NSMM Newsbrief page, which is always updated in near-real-time. Additionally, the nuclear-specific categories which automatically filter the incoming information can be further refined and thus allow the users to access the articles according to more specific geographic or thematic areas of interest (e.g. issues related to specific fuel cycle steps or countries which might not be covered by a more generic newsletter). The NSMM could allow an inspector or analyst to set up filters for their specific area of interest to ensure access to relevant open source information on a near real-time continuous basis.

— Daily monitoring for the generation of a domain-specific newsletter: The information analyst manually reviews all media articles that pass the NSMM filter; selects the articles that are most relevant to the subject (e.g. Nuclear Security); and arranges them in the resulting newsletter according to thematic and geographic area of interest.

NSMM has initially been established in a joint project with IAEA's Division of Information Management (SGIM) with the aim to streamline IAEA's acquisition and analysis of open source information and develop the current information collection production process for the internal news review (SGIM Highlights) to a more efficient system. SGIM has started using NSMM/NewsDesk for the daily monitoring of its 'web sources' (over 150 nuclear-related web sites that were previously monitored manually) in the beginning of 2013. NSMM/NewsDesk is used for selected parts of the production workflow of the SGIM Highlights (see [1]).

As well as the daily news monitoring and newsletter production, SGIM analysts also carry out country-specific monitoring and searches, which typically involves monitoring additional national sources and generates more state-specific information. Following the positive experience in the daily monitoring, SGIM currently evaluates how NSMM/NewsDesk can contribute to country-specific monitoring. A pilot country was identified for the evaluation phase and a list of about 25 national web sites covering governmental, industrial and local media sources is being monitored by NSMM.

NSMM is also used operationally at the JRC for the daily creation of a Nuclear Security News review, which is distributed to selected recipients in the EU institutions. The number of articles going through the JRC workflow on a typical day are as follows: NSMM collects 170000 articles on the monitored sites and filters about 500 articles relevant to nuclear security/safeguards issues. The information analyst selects about 25 articles to be included in the newsletter which is then distributed to selected recipients in the EU institutions. The meta data of the selected articles (including title, description, URL, entity information extracted by NSMM and thematic/geographic tags added by the analyst) are automatically imported to an internal newsletter repository, which provides advanced search and browse functionality to the analyst. As the articles have been reviewed and selected by a domain expert, the repository constitutes a valuable starting point for targeted Open Source studies on safeguards relevant issues. Figure 2 illustrates the information flow for NSMM and the related NewsDesk application as used within JRC.

Developments are ongoing to further improve the functionality of the open source monitoring tools:

— Whereas the nuclear sources used for the daily monitoring focus on English-language sites, country-specific monitoring also requires the collection and analysis of articles in other languages. In this context, the keywords definition that automatically filters the relevant articles will be translated into additional language.

— The current search and browsing functionality of the articles repository requires some prior knowledge of the repository content (what themes and geographic areas are covered). The concepts for visually exploring database content [21] that were developed in the context of trade analysis facilitate the discovery of events and patterns in the database. Some initial tests were carried out and the development of operational tools is under consideration.

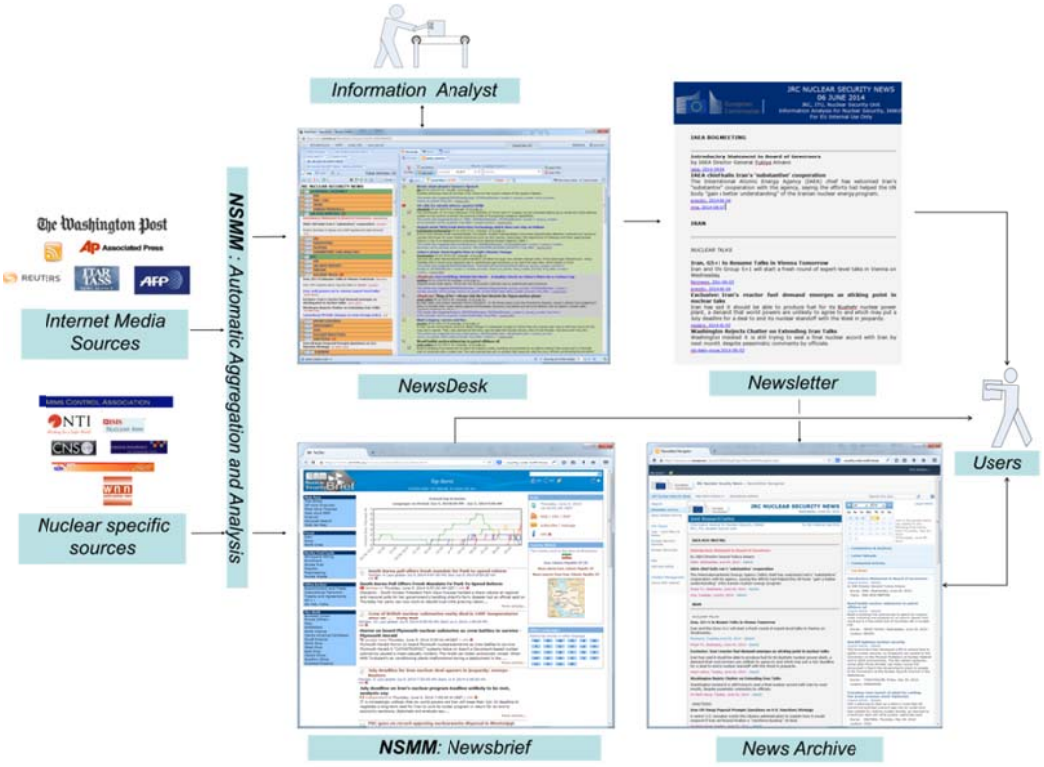

Figure 2. Schematic overview of information flow in a possible setup for open source information collection and analysis based on NSMM and NewsDesk.

4. Conclusions and outlook

As acknowledged by [28]: "The collection and analysis of open source information is an essential element in the State evaluation process. Open source information, including trade and procurement data, can provide early indications of potentially undeclared nuclear activities. Analysis of open-source information is used by the Department to support the State evaluation process, in particular to help verify the completeness".

Open source information analysis can thus play a role in the IAEA by contributing to the establishment of a global state picture within the context of nuclear safeguards. In this paper a particular focus has been put on the possible role and use of trade data for nonproliferation activities, on the open source information monitoring and collection process and on the tools and approaches developed by JRC to support them.

The importance of open source information analysis has the potential of being further increased in the future following the evolution of the state level concept by contributing to the establishment of a global state picture. This might inform the shaping of state specific factors and of the related technical objectives to improve safeguards effectiveness and efficiency.

Acknowledgments

The work here presented is being carried out within the projects Open Source Information for Nuclear Security: applications and tools and Strategic Trade Analysis for Non Proliferation, both funded within the European Commission (EC) Euratom Horizon 2020 Research and Training Programme, as contribution to the EC Support to the IAEA. The Nuclear Security Media Monitor is developed by EC-JRC as collaboration between the Institute for Transuranium Elements and the Institute for the Protection and Security of the Citizen. The authors would like to thanks all the IAEA officers and JRC colleagues that contributed to the activities here presented and provided feedback on the manuscript.

References

[1] COJAZZI G.G.M., et al., "Collection and Analysis of Open Source News for Information Awareness and Early Warning in Nuclear Safeguards". *ESARDA Bulletin No. 50*, December 2013.

[2] BEST, R.A., CUMMING A., "Open Source Intelligence (OSINT): Issues for Congress", Order Code RL34270, Congressional Research Service, 2007.

[3] VERSINO C., et al., *The Big Table. A Software Tool for Nuclear Trade Analysis*. EUR 24823 EN, ISBN 978-92-79-20297-1, JRC62804, 2011.

[4] VERSINO C., et al., *Integrating IAEA's Physical Model with JRC's The Big Table document search tool*. EUR 26215 EN, ISBN 978-92-79-33675-1, JRC77541, 2013.

[5] VERSINO C., et al., *Catalogue of WEB Data Services on Global Trade*. EUR 24657 EN, ISBN 978-92-79-18919-7, JRC62363, 2010.

[6] The Harmonized System by World Customs Organization. Available at: http://www.wcoomd.org/home_hsoverviewboxes_hsharmonizedsystem.htm

[7] United Nations Organisation. COMTRADE consultation site: http://comtrade.un.org/

[8] Eurostat. Easy COMEXT consultation site: http://epp.eurostat.ec.europa.eu/newxtweb/

[9] EU Correlation Table document by DG TAXUD, updated July 2012. Available at: http://trade.ec.europa.eu/doclib/docs/2012/july/tradoc_149781.pdf

[10] Council Regulation No 428/2009 setting up a Community regime for the control of exports, transfer, brokering and transit of dual-use items (Recast). Available at: http://trade.ec.europa.eu/doclib/docs/2009/june/tradoc_143390.pdf

[11] LIU Z., MORSY S. *Development of the Physical Model*. IAEA-SM-367/13/07, IAEA, Vienna, 2007.

[12] Nuclear Suppliers Group. *Guidelines for Nuclear Transfers*, INFCIRC/254/Rev.10/Part 1, 2011. Available at: http://www.iaea.org/Publications/Documents/Infcircs/2011/infcirc254r10p1.pdf

[13] Nuclear Suppliers Group. *Guidelines for Transfers of Nuclear-Related Dual-Use Equipment, Materials, Software and Related Technology*, INFCIRC/254/Rev.8/Part 2, 2010. Available at: http://www.iaea.org/Publications/Documents/Infcircs/2010/infcirc254r8p2.pdf

[14] IAEA, INFCIRC/540 (Corrected). *Model Protocol Additional to the Agreement(S) Between State(S) and the International Atomic Energy Agency for the Application of Safeguards*, 1997.

[15] National Nuclear Security Administration Nonproliferation and International Security Office of Export Control Policy and Cooperation. *A Handbook for the International Atomic Energy Agency's Model Protocol Annex II*, 2004.

[16] National Nuclear Security Administration. *A handbook for the Nuclear Suppliers Group Trigger List Annexes*, NSG Official Use Only, 2002.

[17] The US Department of Energy. *A handbook for the Nuclear Suppliers Group Dual-use Annex*, LA-13131-M, Rev. 1, Official Use Only, 2003.

[18] CUNNINGHAM J.B. *Annex to the letter dated 3 May 2002 from the Deputy Permanent Representative of the United States of America to the United Nations addressed to the President of the Security Council. Goods Review List*. United Nations Security Council, S/2002/515.

[19] The U.S. Department of Energy. *Handbook for Notification of Exports to Iraq ANNEX 3 United Nations Security Council Resolution 1051*, 1998.

[20] HILL, J. *Links between activities identified in Annex I of the Additional Protocol (INFCIRC/540) and items identified in the NSG dual-use list (INFCIRC/254 Part 2).* Cojazzi G.G.M. (Ed.) JRC62398, 2010.

[21] http://www.tableausoftware.com/

[22] VERSINO C., et al., Global Trade Data to Support IAEA Safeguards. *ESARDA Bulletin* (45), 2010.

[23] MARINOVA E., et al., *Analysis of nuclear relevant information on international procurement and industrial activities for safeguards purposes.* Symposium on International Safeguards: Linking Strategy, Implementation and People. IAEA, Vienna, 20-24 October 2014.

[24] ATKINSON, M., VAN DER GOOT, E., "Near real-time information mining in multilingual news". In: *Proceedings of the 18th International World Wide Web Conference (WWW 2009)*, Madrid, Spain, pp. 1153–1154, April 2009.

[25] STEINBERGER, R., et al., "An introduction to the Europe media monitor family of applications". In: Gey, F., Kando, N., Karlgren, J. (eds.) *Proceeding of the SIGIR Workshop on Information Access in a Multilingual World (SIGIR-CLIR2009)*, Boston, USA, July 2009.

[26] http://emm.newsbrief.eu

[27] https://nsmm.emm4u.eu/nucsec

[28] IAEA, *Development and Implementation Support Programme for Nuclear Verification 2014–2015, Vienna, December 2013, STR 377*, Accessed at: http://www.bnl.gov/ispo/docs/pdf/RD%20Programme/D-IS_ProgForNuclearVerific_2014-2015.pdf