



**HAL**  
open science

## Collection and evaluation of open source information relevant to the Additional Protocol

Lucie Millot

► **To cite this version:**

Lucie Millot. Collection and evaluation of open source information relevant to the Additional Protocol. INMM 2019 - 60th Annual meeting, INMM, Jul 2019, Palm Desert, United States. irsn-03881173

**HAL Id: irsn-03881173**

**<https://hal-irsnn.archives-ouvertes.fr/irsnn-03881173>**

Submitted on 1 Dec 2022

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Copyright

## **Collection and evaluation of open source information relevant to the Additional Protocol**

Lucie MILLOT

Institut de Radioprotection et de Sûreté Nucléaire (IRSN)  
Non-Proliferation and Nuclear Material Accountancy Department  
International Safeguards Unit  
B.P. 17 – 92262 Fontenay-aux-Roses Cedex, France

### **Abstract:**

Open source data mining is one of the most explored topics nowadays to discover the information and knowledge from the existing data. Its use is encouraged by IAEA guidelines “Safeguards Implementation Practices Guide on Provision of Information to the IAEA”.

The complexity of the analytical work to ensure the completeness of the declaration is to define on the one hand if all covered entities have been targeted and in the other hand if all activities have been declared.

Through the French example, this work describes how open-source information data mining allows initializing and updating the entities list, validating the addresses declared in PASTEL, the French custom portal created to enhance Additional Protocol (AP) declarations, and identifying new research programs.

In order to support IAEA, France has introduced methodologies using open source information collection and evaluation to improve French Additional Protocol declaration and insure the completeness and correctness of information declared to IAEA. Indeed, French Additional Protocol is focused on various activities (research on fuel cycle, dual use-good trade, mining and waste transfer), with an important amount of data (addresses, programs, new registrants etc.) to be processed.

This paper identifies two methods to improve the detection according to different AP's articles. Indeed, this study shows that the use of company directories such as Kompass® improves the declaration for AP's article for the manufacture and export of dual-use goods, even if the number of lines declared remains stable. A cross matching with dual-use license is under studying and looks promising.

It also indicates that the systematization of the looking for research programs increases the number of programs declared to the IAEA but also improves the new registrant's detection.

### **Introduction**

Since May 1997, the Board of Governors of the International Atomic Energy Agency (IAEA) approved the text of Model Protocol Additional (Model AP) to the Agreements between States and the IAEA for the application of safeguards (Model Protocol) INFCIRC/540 [1]. Regarding France, the French additional protocol (French AP) [2] to its safeguards agreement [3] has entered into the force the 30<sup>th</sup> of April 2004.

To collect information for the AP, the IAEA recommends to the State or Regional Authority (SRA) to conduct an open source search initially and periodically during updates [4].

In this paper, open source information is defined as information that neither classified nor proprietary, these include media sources, social media, scientific literature, etc. Most of these data are freely available on the Internet.

France has introduced methodologies using open source information collection and evaluation to improve French Additional Protocol declaration and ensured the completeness and correctness of information declared to IAEA.

Within its mandates given by the French authority, the Euratom Technical Committee (CTE), the Non-Proliferation and Nuclear Material Accountancy Department of the French Institute for Radiological Protection and Nuclear Safety (IRSN) decided to communicate easily with registrants (stakeholders, universities, researchers, etc.) by using of forms (noted Form X) instead of the number of the article to carry out the AP at state level. The table below gives the transcription between French AP, the Model AP and the forms. Following discussion in this paper will focus on forms.

	Model AP [1]	French AP [2]
Form A	2.a.(i)	2.a.(i)
Form B	2.b.(i)	2.b
Form C	2.a.(x)	2.a.(viii)
Form D	2.a.(viii)	2.a.(vi)
Form E	2.a.(iv)	2.a.(iii)
Form T	2.a.(ix)	2.a.(vii)

**Table 1: Comparison of the articles of the Model Additional Protocol and the French one.**

## **I. How could open source data be used to improve the Additional Protocol?**

The issue of the analytical work is to ensure the completeness of the declaration. The SRA have to prove that all covered entities have been targeted and all activities have been declared.

As Cairn [5] and Rockwood [6] said, there are many uncertainties about the information required to declare and stakeholders who are familiar with technical fields are not familiar with AP reporting requirements.

To implement the French AP, IRSN identified three ways to use open source: to check addresses declared by the registrants, to detect new registrants and to find new research programs.

### **a. Addresses verification**

One of the IRSN tasks for this declaration is to check that the new addresses filled in PASTEL, the web portal dedicated to AP declaration, are correct and do not already exist in the database. This addresses verification is an issue especially in case of complementary access.

Therefore, for each address IRSN have to:

- 1) Check on internet if:

- a. the addresses exist physically (use of satellite imagery for example)?
  - b. The name of the company is an acronym and what this acronym means in French, or in other languages?
  - c. the address is complete?
- 2) Check that the address does not already exist in PASTEL.

Finally, there are two possible outcomes: the address is a “true” new address, and in that case it is added in the database, or the address already exists in PASTEL. For information, each year, around 35 new addresses are added.

This allows for IRSN and IAEA to ensure the completeness of the declaration. From the IRSN point of view, this verification insures that all French registrants are targeted and declare, especially in the case of start-up companies which are acquired by biggest companies.

## **b. New registrants detection**

### *i. Initialising the entities list*

In the light of the French experience to initialize the French entities list, IRSN recommends the use of company directories with a governmental information cross matching. Indeed, company directories can give much information about the trade analysis [7].

The French entities list was created in 2003 with 4000 pre-selected entities [8]. The working method was essentially based on information cross-checking with company directories (KOMPASS© 80%), governmental information (15% of entities) and other open source data (5%).

This is the stepwise approach carried out to the first census:

1. request on key words used in the nuclear industry as “nuclear”, “radioactive”, “atomic” or “uranium”;
2. request on R&D field key words;
3. request using annex I criterion;
4. request using annex II criterion;
5. remove from the list duplicates.

After this census, IRSN had compared his own data with governmental information and completed them with other open source.

### *ii. Updating the entities list*

The entity list is updated every year and has now 216 French entities. Each year, IRSN added 10.8 entities from the list and remove 5.5 entities [9].

This paper is focusing on the number of new registrants over a five-year period, between 2012 and 2017, as it is presented Figure 1.

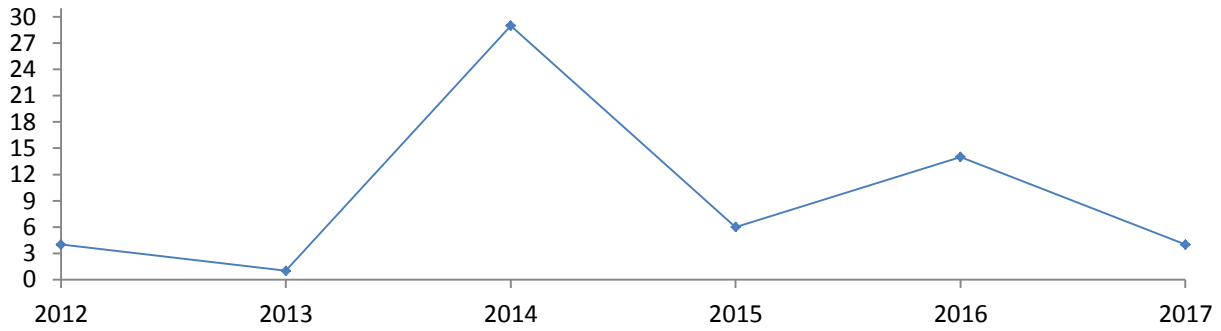


Figure 1: Number of new registrants from 2012 to 2017.

In 2014, IRSN did a general census and added 28 new registrants. The same method that the one presented above was used. Since 2015, an open source mining is systematically done, each year, using different kinds of sources as it is presented in Figure 2.

On the bar chart (Figure 2), the green bars indicate the diverse open data used: Press, Company directories as KOMPASS®, CORDIS, and Networks.

CORDIS is web site which is the European Commission's primary portal for results of EU-funded research projects; this website presents all research programs and shows the funds provided by companies.

Network websites are defined in this work as all web sites where a link with other entities can be done. This includes social media like LinkedIn®, but also nuclear association as AIFEN or student website like CampusFrance for PhD on nuclear fuel cycle. In purple the others sources are mostly data given by registrants already targeted.

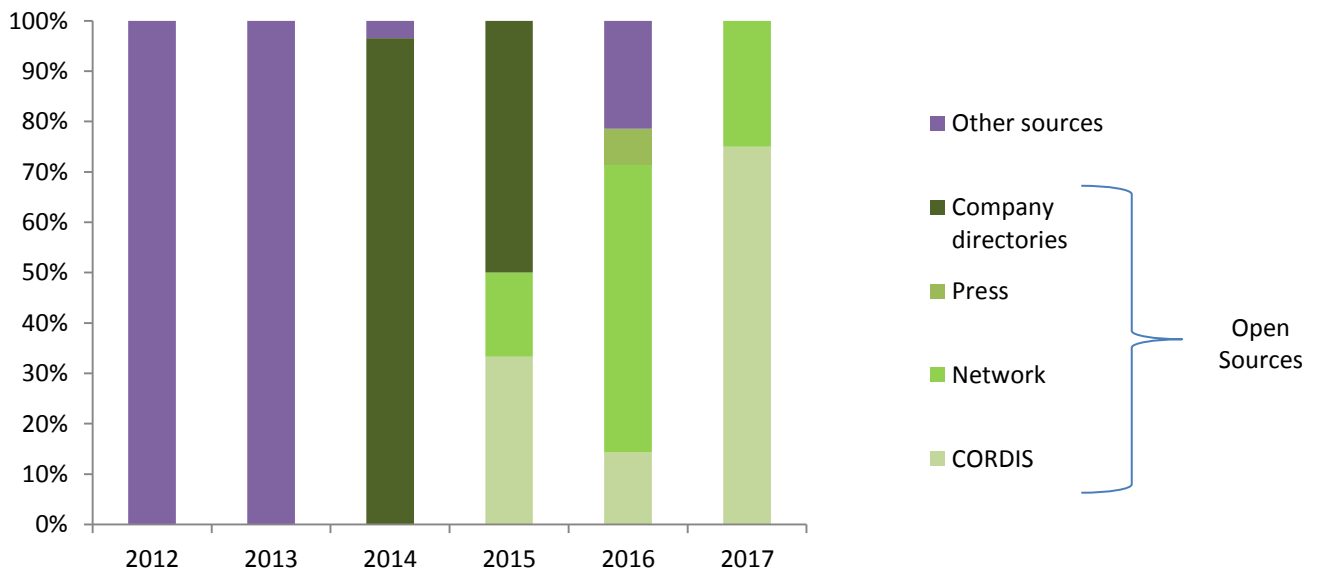


Figure 2: Distribution of new registrants added between 2012 and 2017.

IRSN found a significant part of new registrants with Company directories, with 31 new registrants added during this period. The second significant contribution was Network. While the impact of using CORDIS to find new registrants was less significant about numerous (6 new entities added in five years) its use is really useful and efficient especially because the mining is less time-consuming.

### c. Identifying new research programs

Since 2015, a systematic study of the research programs was implemented to ensure the completeness of the French declaration (form A). To do this, IRSN focusses essentially four types of sources: CORDIS, LinkedIn®, the press and the scientific literature.

Although most of the reported programs were not identified in this research, between 15% and 50% of new programs detected were through open source (see Figure 3). Actually, the other new research programs are mostly PhD projects or bilateral research agreements declared by stakeholders. For example, in 2016, IRSN added doctoral schools where, regardless of their home laboratory, students must be registered. All of these schools were found through open-sources and gave a significant number of new programs declared to IAEA (in red in figure 3).

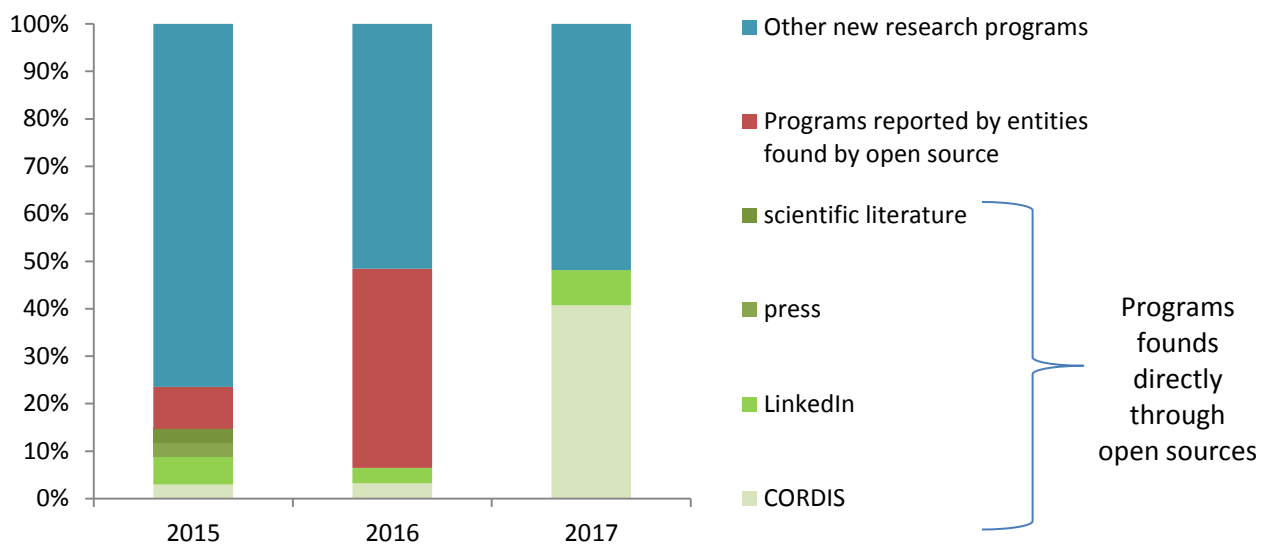


Figure 3: Distribution of new programs by source of data between 2015 and 2017.

This paper shows from 10 and 50% of new reporting programs was found directly through open source data. A significant part of new programs are found on CORDIS, around 4-5 by year. The second contribution to this census is LinkedIn® with an average of 1.67 new programs found with this social media by year. Press and scientific literature give few results and is time-consuming.

A closer look indicates that the impact of open source is larger. Indeed, a snowball effect was observed: new programs found with open source allow IRSN to find new entities which will declare new programs. For example, if a new program is found (mostly in CORDIS), IRSN will check all the French participants to be sure that PASTEL data base is updated. From this mining, IRSN will add between 2 and 3 new registrants which will declare also new programs, not found by open source data mining.

Hence, the use of open source directly and indirectly had an important impact on Form A declaration (see Table 2). Actually, for the reporting period 2016 and 2017, 50% of new research programs reported to IAEA were found directly or indirectly through open source data mining.

	Program directly found with open source directly (%)	Program reported by entities found by open source (%)
2015	15%	8%
2016	10%	40%
2017	50%	0%

Table 2: Distribution by year of programs found directly or indirectly by open source mining.

## II. Impact on declarations

IRSN studied the impact of the new methods to census on French AP declaration. The results are given in the Figure 4 which shows the correlation between the total number of French AP registrations and the number of new registrants over a five-year period.

The first observation is that the number of Form E declarations sent to IAEA remains stable over this period (in green in the Figure 4). Measure the impact of the new registrants is complex because manufacturing of items under the AP requirements is correlated with the market condition. Indeed, some stakeholder which has activities targeted one year could not has the following year.

Nevertheless, it notes despite the ad of new registrants that the impact on declaration sent to IAEA for this form is not significant. This suggests that the initializing of the list had been efficient. However, a cross-matching with dual-use licenses is under studying and looks promising to insure of the completeness of Form E and T declarations [10].

This study shows that the number of programs (Form A) in blue in the Figure 4 and projects (Form C) in red in the Figure 4, reported to IAEA rise significantly since 2015. IRSN considers that the number of reported programs depends on program funding, as Horizon 2020 which is the funding program created by the European Union/European Commission to support and foster research in the European Research Area but also on new registrants contacted. Indeed, this rise in declaration is due to the snow ball effect presented above.

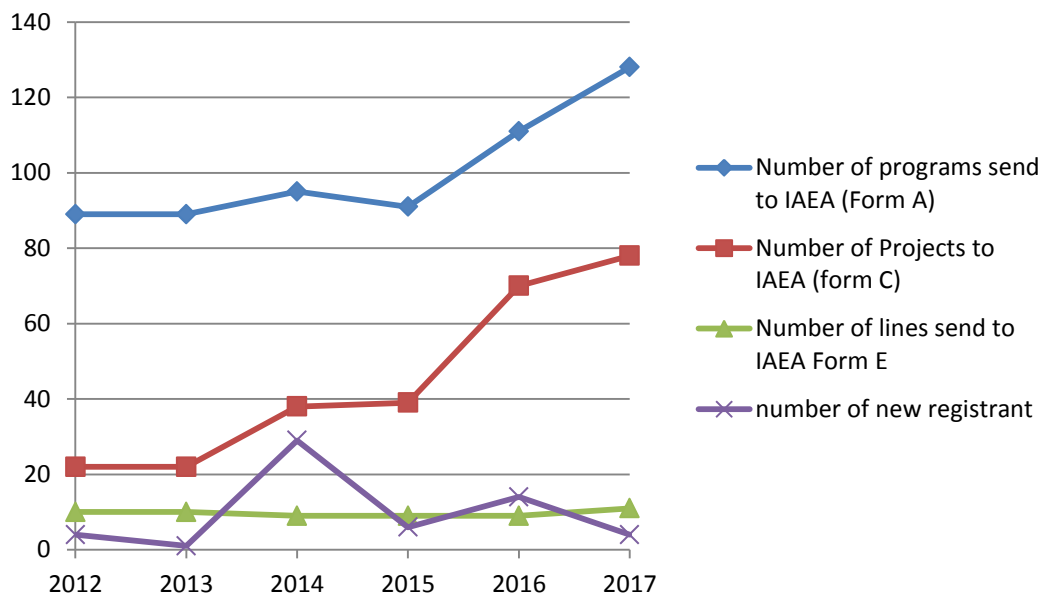


Figure 4: Evolution of the number of reported programs, projects, number of lines sends to Form E and the number of new registrants over a six years period.

## **Conclusion**

France has used open sources data mining for three purposes: checking addresses declared by the registrants, detecting new registrants and identifying new research programs.

This paper identified two methods to improve the detection according to different AP's articles:

- 1) Using company directories to initialize the entities. This census is efficient, and provides a long-term stability for form E declaration. However, a cross-matching with dual-use licenses is under studying and looks promising.
- 2) Using of social media as LinkedIn© and CORDIS for detecting both new entities and new programs. This method highly impact Form A and Form C because systematize the data mining on research programs (form A) or projects (form C) increases the number of programs declared to the IAEA and improves in the meantime the new registrant's mining. Actually, open source data-mining detected directly or indirectly 50% of programs reported to IAEA.

These results instil the confidence of the completeness and correctness of information declared to IAEA.

However this open sources data-mining is time-consuming, about 20 men-days every year in average. The use of machine learning to improve this data-mining and save time is clearly a future area for debate.

## **References**

[1] Model Protocol Additional to the Agreement(s) Between State(s) and the International Atomic Energy Agency for the Application of Safeguards, IAEA, Vienna (1997).

[2] The Protocol Additional to the Agreement between France, the European Atomic Energy Community and the International Atomic Energy Agency for the Application of Safeguards in France, IAEA, Vienna (2005).

[3] The Text of the Agreement of 27 July 1978 Between France, the European Atomic Energy Community and the International Atomic Energy Agency for the Application of Safeguards in France, IAEA, Vienna (1978).

[4] Safeguards implementation practices guide on provision of information to the IAEA, IAEA, Vienna (2016).

[5] CAIN, R., "DOE Additional Protocol Implementation", presented at the NGS Summer Seminar Series, 2011.

[6] ROCKWOOD, L., Evaluation of the Impact of the Model Additional Protocol on Non-Nuclear-Weapon States with Comprehensive Safeguards Agreements, Swedish Radiation Safety Authority, Vienna (2018).

[7] COJAZZI, G., VERSINO, C., WOLFART, E., RENDA, G., JANSSENS, W., Tools for trade analysis and open source information monitoring for nonproliferation, IAEA Symp. Int. Safeguards (2014) 318.



[8] SEVIN, J., DANIEL, G., “French Additional Protocol: 10 Years of Implementation”, Symposium on International Safeguards Linking Strategy, Implementation and People presented at the, Vienna, 2014.

[9] MILLOT, L., “Benefits of using a Declarative Web Portal to make the Additional Protocol Declarations- Application with the French Portal PASTEL”, IAEA Symposium on International Safeguards presented at the, Vienna, 2018.

[10] VIAL, E., OLIVIER, L., Export Control and Additional Protocol Working in Synergy to Strengthen Nuclear Non-Proliferation: Implementation in France, Proceedings of the ESARDA 41st Annual Meeting (2019)