EMM ontology on GMOs

Customization of MedISys for the monitoring of GMOs without positive safety assessment

RIKILT Report 2012.016

T.W. Prins, J. Top, E.J. Kok and H.J.P. Marvin

EMM ontology on GMOs

Customization of MedISys for the monitoring of GMOs without positive safety assessment

T.W. Prins, J. Top[1], E.J. Kok and H.J.P. Marvin

Report 2012.016 **November 2012**

Project number: 121.7280701
BAS-code: BO-08-011.02-023
Project title: Voorstudie EMM ontologie

Project leader: H.J.P. Marvin

RIKILT Wageningen UR
(University & Research centre)
Akkermaalsbos 2, 6708 WB Wageningen, The Netherlands
P.O. Box 230, 6700 AE Wageningen, The Netherlands
Tel. +31 317 480 256
Internet: www.wageningenUR.nl/en/rikilt

[1] **Wageningen UR Food & Biobased Research, Fresh Food and Chains**
Bornse Weilanden 9
6708 WG Wageningen

The research was conducted in 2011 as part of the Dutch 'Beleidsondersteunend Onderzoek (BO) programma, domein Voedsel, Dier en Consument (VDC), thema Voedselveiligheid Nieuwe Technologieën (BO-08-011.02-023)'.

Distribution list:

- Ministry of Economic Affairs, Department 'Plantaardige Agroketens en Voedselkwaliteit' (EZ-PAV); B.J.M. Crijns, J.L. Thio.
- Ministry of Economic Affairs, Department 'Directie AgroKennis' (EZ-DAK); R.A. Donker.

Summary

Within the European Union, genetically modified organisms (GMOs) are allowed on the market only after assessment and approval of their food/feed and environmental safety according to the European regulation 1829/2003. Most GMOs are developed and produced outside Europe and these may enter the European market without having been authorised in Europe. To be able to identify such unauthorised GMOs it is important to be aware of their existence. This knowledge may be used to develop customised laboratory detection methods that will guarantee the detection and identification of these GMOs at the border of the European Union (EU).

A diversity of repositories and documents published on the WWW provides potential sources of information for new GMOs that may have safety issues linked to them. However, present search strategies are not yet optimal.

The Europe Media Monitor (EMM) is a news gathering engine which is operated by the Joint Research Centre (JRC). EMM analyses news items published on the WWW, extracts relevant information, aggregates the collected information, issues alerts, and produces visual presentations of the information collected. Within the framework of the present study, we have analysed the potential of EMM to find and collect information on the WWW on GMOs that have not yet been assessed for their food/feed and environmental safety. Furthermore, we have analysed the current EMM GMO filters to see whether it is possible to improve their performance.

We conclude that EMM, as developed by the JRC, provides a powerful tool to facilitate a dedicated search for GMOs like plants, animals and micro-organisms that have not (yet) been assessed for their safety. Important prerequisites for finding relevant publications are the careful selection of terms that are either included or excluded in the search system. The present study proposes alternative sets of keywords to improve the quality of the query results. In addition to testing alternative sets of keywords to reduce the numbers of irrelevant hits on the basis of the basic query, i.e. the identification of unapproved GMOs, we propose to use ontologies with respect to the GMO domain to improve the result quality of the query. In addition, selecting proper web sources for this specific purpose and ranking hits on the basis of the ontology will increase both recall and precision of the resulting findings.

EMM is currently equipped with a tab from EFSA, containing different filters for different GMO- related queries. For the intended use as described in this report an extension of the filters seems advisable. We hope to contribute to EFSA's mandate for 'Emerging Risks' with our suggestions to track and identify the GMOs without a food/feed or environmental safety assessment as part of an early detection system for newly identified hazards.

Contents

1 Introduction

1.1 Problem statement

For the enforcement of European GMO regulations, it is important to be aware of new GMOs (Genetically Modified Organisms) that are being developed elsewhere in the world that do not have undergone a safety assessment. These GMOs should be prevented to enter the European market since they are not in compliance with the European regulation and therefore may cause harm to humans, animals and the environment. Early identification of new GMOs that may present potential risks may allow the development of customised detection methods that will guarantee the detection and identification of these GMOs at the border of the European Union (EU).
A diversity of repositories and documents published on the WWW could provide sources of information for detecting new GMOs. However, present search strategies are not yet optimal.

1.2 Safety assessment

In the EU, regulations are in place (Regulation (EC) No 1829/2003 and 1830/2003 (EC 2003)) that prohibit the entry into the EU of unauthorised GMOs. Part of the approval procedure for new GMOs is a safety assessment that investigates the safety of the GMO with respect to human food, animal feed and environmental aspects (EFSA 2011). Only those GMOs that have been assessed for their food/feed safety and are demonstrable at least as safe as their conventional counter-parts, are allowed into the European market. For the cultivation of new GMOs within the European a similar approach is followed with relation to the introduction into the European environment.

1.3 Background

In practice, GMOs that are not yet approved in the EU, may already have a positive safety assessment in other parts of the world (asynchronously authorised GMOs). These GMOs are not allowed in the EU, but in general their safety considerations may be less pronounced, at least in terms of food/feed safety assessment.

More relevant from a safety point-of-view are the GMOs that do not as yet have a positive safety assessment in any part of the world. It is this group of GMOs that demands special attention of control bodies. The enforcement laboratories require knowledge on the potential presence of GMOs not having a positive safety assessment in order to be able to develop appropriate detection and identification strategies.

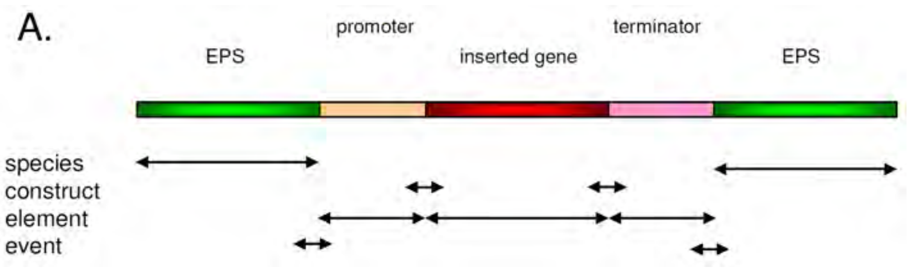


Figure 1. Simplified inserted sequence of a genetically modified plant. EPS = endogenous plant sequence. Taken from Prins et al. (2008).

In short, for example in case of a GMO plant, the plant is typically genetically modified by the addition of one or more genes in order to give this plant a benefit like herbicide resistance, insect resistance, abiotic stress tolerance (See Figure 1). GMO detection can be based on a method that identifies the GMO itself by an event-specific method, or on the presence of certain elements (promoter-gene-terminator) that are potentially present in a GMO. Once the presence of different elements that have been detected cannot be explained by an existing GMO of which the elements are known, it is indicative that the GMO present in a food or feed sample is unknown and therefore unauthorised. This was for example demonstrated by the detection of illegal rice like Bt63 (EC 2008; Grohmann and Made 2009). Some other incidents have been summarised by Ruttink et al. (2010b).

Some examples of unapproved GMOs have been described in the scientific literature. In the example of Coban (Ruttink et al. 2010a; Ruttink et al. 2010b), a plant preparation containing *Arabidopsis thaliana* that was genetically modified by adding a human intrinsic factor (rhIF), was found in a B_{12} preparation. The authors traced back the information that could be found on the internet ('documentation space') about the Coban product ('physical space') and showed that knowledge-technology-based use of the documentation space could indeed provide information that is independent of laboratory analyses. Although this example is very enlightening, the predictive approach may be more valuable than this retrospective approach.

A method that would alert controlling bodies about the presence of new GMOs such as described above will help these organisations to develop dedicated laboratory testing methods to identify the appearance of these GMOs in the food/ feed production system.

1.4 Finding GMOs without positive safety assessment

Tools that support automated search of articles from internet are judged by their 'recall' and 'precision' with respect to the ideal set of hits. This ideal set of hits ('the golden standard') contains all useful hits and nothing else. 'Recall' measures how many of the ideal hits have been found, and 'precision' expresses how many non-relevant hits have been found. Improving recall usually goes at the expense of precision and vice versa; a very strict search query produces little garbage but possibly excludes good results, while a loose query generally also produces lots of irrelevant hits. An overall measure for success is the f-measure, a weighted average between recall and precision.

In Europe, such automated search engine has been developed (e.g., Europe Media Monitor (EMM)), and this system may provide an excellent platform to search the internet for unauthorised GMOs.

This report describes our analysis of the potential of EMM as a tool to find articles published on the WWW that are related to the release of new unauthorised GMOs. In the following chapters, different aspects and applications of EMM will be discussed.

2 Europe Media Monitor search tool

In Europe, the Joint Research Centre (JRC) has produced software, the Europe Media Monitor, which collects reports from news portals world-wide in 43 languages. EMM is a news gathering engine that classifies articles, analyses the news and extracts information, aggregates the information, issues alerts, and produces visual presentations of the information (http://emm.newsbrief.eu/overview.html). EMM applies text mining techniques to screen different types of media on the World Wide Web, like websites, databases, blogs etc., using specific search queries. A query consists of a logical combination of search terms (using Boolean operators such as AND, OR and NOT), which is then matched to the available sources. This is similar to advanced use of Google, although EMM uses dedicated Natural Language Processing technology to improve the pattern matching process. Using this approach, EMM can select news items are potentially relevant, although the item itself does contains the information explicitly (serendipity). The selection of search terms and their logical combination is a critical component of EMM.

Within EMM, MedISys (Medical Information System) is a 24/7 news alert system that was adapted for medical and health-related topics. MedISys is run by the Health Threats Unit at DG Health and Consumer Affairs of the EC, in collaboration with JRC in Ispra, Italy. The purpose is to strengthen the network for surveillance of communicable diseases and the early detection of bioterrorism activities. Online information sources are used to rapidly detect, track and assess threats to enable providing early warnings (https://medisys.jrc.it/medisys/aboutMedISys.html). Currently, one field within the EMM-MedISys focuses on general food safety issues, but not yet on the application of biotechnology for development of genetically modified plants, animals and microorganisms (EMM-MedISys: Figure 2).

Figure 2. General search for GMO in EMM MedISys (date accessed: August 11, 2011).

Recently, the European Food Safety Authority (EFSA) reached a service level agreement with DG-JRC to customise the Medical Information System for the monitoring of reports published on food and feed hazards (Linge and Belyaeva 2011). Within this framework 139 categories and 211 filters were created. However, only 5 filters are related to GMOs. JRC has introduced a new EFSA tab in the menu structure of the restricted MedISys site: https://medisys.jrc.it/medisys/categoryedition/efsa/en/efsa.html. In the GMO tab, the following five different filters are present (Figure 3):

1. Food and feed safety
2. GM animals
3. GMO environment
4. GM microorganisms
5. Molecular characterisation

These five filters have been designed to retrieve media information on GMO-related issues in general. The approach for detecting GMO-related information proposed in the current report focuses specifically on those GM organisms that do not have a positive safety assessment. This could become part of the tab for monitoring GMOs, for example as a deeper filter in 'Food and Feed Safety'.

Customization of MedISys for the monitoring of food and feed hazards

Project developed on the procurement project SLA/EFSA-JRC/2010/EMRISK /01 (Linge and Belyaeva 2011)

Within the Service-level agreement (SLA), EFSA and JRC have extended the threat detection system MedISys to food & feed hazards. The media coverage of MedISys has been extended by 300 sources (task 1a of SLA). Over 200 filters for common food & feed hazards have been added covering additives and supplements, animal health, biological hazards, contaminants, feed, food contact material, GMO, nutrition and allergens, pesticides and plant health (task 1b). Within these filters, areas of broader scope were covered (i.e. drivers of change which could have an indirect impact on the food chain). The multi-lingual filters have been tested and fine-tuned to reduce the volume of news articles per day to a manageable level; users can select official or general news sources (task 1c). Deduplication (task 2f) is performed with the same algorithm used for Europe Media Monitor (EMM).

The menu structure of the web site has been amended so that all filters are easily accessible via the web interface. EFSA has its own tab on the menu of the restricted MedISys site which allows all EFSA staff to screen news articles for the new filters (tasks 2d-e). Press coverage on EFSA is available for press officers and senior management.

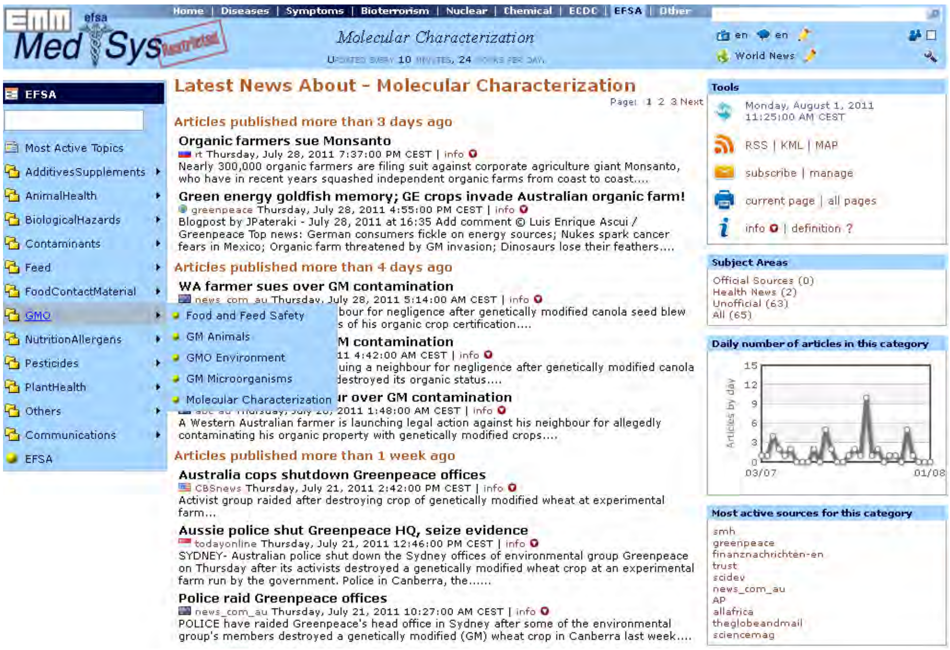


Figure 3. Search for GMO-Molecular characterisation in EMM MedISys, dedicated to EFSA Food and Feed safety (date accessed: August 11, 2011).

2.1 Improving EMM for a customised search for GMOs without positive safety assessment

We propose to implement a dedicated search strategy for unapproved GMOs within the existing EMM GM-filters. An automated search (24/7, with news flashes every 10 minutes) would facilitate a dedicated search for GMOs that are reported somewhere on the WWW but have not been noticed before by regulation bodies. These GMOs may not have a positive safety assessment. Currently, experts and risk assessors interested in these categories perform incidental searches by browsing the internet manually. They may use common web browsers or literature-related sites like Web of Science (http://apps.webofknowledge.com). Some advantages of a manual search are:

- The specific knowledge that a researcher has allows for an interactive, in-depth search strategy (the ability to read and interpret the set of hits to obtain additional search criteria that reduce or extend the present set);
- The ability to search for and analyse complex figures and pictures that do not show up as a hit because the text is not indexed;
- Detailed knowledge of the domain and concepts and terms used. Also the analysis of the output generated by EMM requires specialised knowledge. A researcher needs to elucidate relevance of the hits found and how they relate;
- The ability to switch between sources when considered favourable.

Major disadvantages of a manual search approach are:

- The limitation to the set of languages that the researcher in question masters;
- The limited (incomplete) set of keywords that will be used;
- The time limitation for such an elaborate screening, in terms of frequency as well as time per search.

To overcome the limitations of manual search, an automated search by means of a dedicated Europe Media Monitor is preferable. Such a tool will increase the amount of hits while reducing the time needed from experts. In the following chapters, the automated process will be discussed. We propose to complement and improve the existing EMM GMO filter to the specific use of tracing those GMOs that do not have a positive safety assessment. Different search strategies will be discussed that lead to an optimized list of 'hits' on the WWW and dedicated databases.

3 Requirements for defining the desired search outcome in EMM

To retrieve information specifically on GMOs that have no or a negative safety assessment, it is important to clearly define the search strategy that EMM should be executing. Here, the GMOs to be addressed, as well as the search criteria (ontology) are important to define:

a. Which GMOs are targeted and which are to be excluded?
b. Which sources or internet sites are interesting to include in the search?
c. How can the process of creating search queries be improved to increase the chance on finding GMOs with no or a negative safety assessment?

3.1 Which GMOs are targeted?

In Europe 50 GM plants have been approved (status October 2012). The approval of 5 GM plants has expired and their approval status is now subject to withdrawal. No GM animals have been approved for food and/or animal feed use. Regarding GM microorganisms, only bacterial biomass (pCABL) and Yeast biomass (pMT742 or pAK729) have been approved for use in animal feed (EC).

Authorised GM plants are easy to define in an EMM search since they are restricted to the 38 approved GM plants. All GM animals are targets in the search strategy, since no GM animal is approved for market anywhere in the world. The identification of GM animals for which a positive safety assessment has been performed in other parts of the world needs to be investigated. GM microorganisms (GMMs) are a miscellaneous field. Many GMMs are approved for the production of enzymes and secondary metabolites, but only under contained use. These GMMs are often not applied in human food or animal feed, with the exception of yeast cells that produce rennin. Further to this, products may enter the EU from non EU countries were GMMs may already have been approved. Although GMMs that produce medicine do not fall under the responsibility of EFSA, we propose to include them in the current search strategy since they may end up in the food chain. Another group of organisms that is gaining attention is that of algae. Algae, that are neither plants nor microorganisms, can be produced as a source for biofuel, generating a large biomass as by-product. Although these algae are not primarily suitable for feed, they can be used as such and therefore it is recommended to include GM algae in the EMM search.

In general, the focus of this study is on GM species that have no positive safety assessment in Europe or elsewhere in the world, specifically on plants (including algae), animals and microorganisms. To be relevant in regard to human or environmental risk, the GMO should be in an advanced state of application, i.e. production should be at such a level that it can be expected to be present in the food or feed chain. GMOs can be divided into different classes as depicted below:

EU-Authorised GMOs:

I) Admitted in the EU (with positive safety assessment).

EU-Unauthorised GMOs:

For EMM, this is regarded as one group, but it can be subdivided into the following categories:

IIa) Admitted in countries outside the EU (with positive safety assessment*)

IIb) Unauthorised inside or outside the EU, with positive safety assessment*

IIIb) Unauthorised inside or outside the EU, without positive safety assessment*

1. Known GMOs
2. Unknown GMOs that contain known GMO elements
3. Unknown GMOs that contain unknown GMO elements

GMOs in class I) class IIa) and class IIb) have a positive safety assessment and are not of interest for this study. Hence, these classes should be ruled-out from the search results. Also those GMOs that are in the pipeline to be admitted in the EU (EFSA 2012) are to be excluded. See Annex I (EU Authorised GMOs).

* A clear definition should be set as to what requirements a correct and complete safety assessment should meet up to. Only then the list of countries that show similarity to the EU market approval system can be generated. Currently, the data used in Annex I was derived from the official EU list of approved GMOs, but can be extended to the list from CERA (CERA), describing approved GMO plants in Argentina, Brazil, Burkina Faso, Canada, China, Colombia, Czech Republic, El Salvador, European Union, India, Japan, Korea, Mexico, Paraguay, Philippines, Russia, South Africa, Switzerland, Taiwan, United States and Uruguay. Additional approved GM plants are presented by ISAAA (ISAAA) and comprise the countries Australia, Bolivia, Chile, Costa Rica, Egypt, Arab Rep., Honduras, Iran, Islamic Rep., Malaysia, Myanmar, New Zealand, Pakistan, Romania, Singapore, Sweden, Thailand, and Turkey. Only a few of these approvals are restricted to one country. The majority has been approved by two or more countries. The accuracy and completeness of the authorization of GM plants was not checked in detail for every country in this stage. Besides the approved GMOs described above, it is even more important to list those GMOs that have received a positive safety assessment. On the ICGEB site, a Risk Assessment Search Mechanism (RASM: http://rasm.icgeb.org) provides an extensive database on the approved GMO plants world-wide. This database will be used as reference for GMOs with a positive safety assessment.

In the list of GM plants, crops like carnation, petunia and creeping bentgrass, and animals like glowfish, are obviously not intended as food or feed. Therefore, it can be expected that they have no or a negative food/feed safety assessment. Nevertheless, they will be included in the search list.

GMOs in class IIIb1 are known GMOs with no or negative safety assessment and can be searched for by GMO name since this information is available.

GMOs in class IIIb2 are unknown GMOs with no or negative safety assessment, but with known GM elements. These GMOs can be found by searching for presence of elements that are known or expected to be in unknown GMOs. Some examples are the application of *Mir*, coding for a sweet tasting Miraculin compound (Hiwasa-Tanase et al. 2011), *vip3*-related genes as an alternative for *cry*-like genes (Singh et al. 2008), cowpea trypsin inhibitor *CpTi (Su et al. 2011)*, aryloxyalkanoate dioxygenase *aad-12* (Wright et al. 2010), Amaranth Albumin *1AmA1* (Tamas et al. 2009), and *Galanthus nivalis* agglutinin *gna* (Ewen and Pusztai 1999). *Pinellia ternata* agglutinin *pta* (Yao et al. 2003) etc. See Annex II (UGM elements). The list will be extended using the controlled vocabulary in the GMO database Euginius that is currently being developed by BVL (Berlin) and RIKILT. Euginius is a transnational cooperation with the intention to make this the European GMO database. Euginius is expected to be online in 2012 and will be updated on a regular basis. Therefore, Euginius is a valuable source of relevant keywords for unauthorised GMOs. Elements that frequently occur with unauthorised GMOs can be selected for to be included in the list of UGM elements described in Annex II. In return, Euginius can show the RSS-feed of the EMM on its homepage, while relevant output of EMM can be included in the Euginius database.

GMOs in class IIIb3 are very hard to identify. The GMOs are unknown and have therefore no unique identifier and no information is available as to what elements are present. One approach is to identify these GMOs by generic keywords, and by the common name and Latin name of the species (See Annex III) in combination with GM. Detecting this class of GMOs on the web is a major challenge.

3.2 Which areas or internet sites are interesting to include in the search for GMOs with no or negative safety assessment?

Several sources can be identified in which this group of GMOs may be reported. Scientists can publish results in a (peer reviewed) scientific article, or as a popular publication on the internet. Governmental organisations can report the existence of such GMOs either as an advertisement or as a warning. Non-governmental organisations (NGOs) can report their existence in case such a GMO is found. Knowledge about useful sources (specific websites) is a crucial factor when searching for this type of relatively informal data. Linge and Belyaeva (2011) provided three lists of data sources (AFCWG, DATAQUEST and PRASSIS), of which the sources from PRASSIS only come from official organisations. When addressing unauthorised GMOs with no or a negative safety assessment, it may be pivotal to include all available sources instead of limiting governmental organisations since GMOs may have additional information available in their network. See Annex IV (Internet sites). The list is sub-divided in literature sources, Governmental sources, NGO sources and general internet links.

3.3 How to build optimized search queries to increase the chance on finding GMOs with no or a negative safety assessment?

Under section 2.1, several lists of terms have been defined that should be included in a search for GMOs with no or a negative safety assessment. In a separate list (See Annex V: Generic keywords), also generic terms have been defined that aid the search for this type of GMOs. Keywords that should be excluded are those GMOs that have a history of safe use, and also have a positive safety assessment. These keywords can be compiled by defining terms that create noise in the search results. One major disadvantage of this approach is that in the case of stacked GMOs, where one of the events is already approved, the stack will not show up as a possible GMO without positive safety assessment. A possible solution is to override the exclusion in case words like [stack, stacked, combined, stacking, multiple etc.] are present. Linge and Belyaeva (2011) already defined a list of Exclusion terms (Annex 3. Exclusion terms (list 3) in their report.

Besides the sole presence of keywords, the links between keywords is also relevant. When two terms that are searched for occur in the same text line, or even adjacent to each other, the predicted value should be higher than when two words are mentioned by the same source, but without any relationship. This should be taken into account when ranking the hits.

4 Possible adaptation of the existing EFSA MedISys GMO filter

As stated in the introduction, EFSA, in collaboration with JRC, has recently adapted the MedISys for GMO analysis. The search for GM plants is already embedded in the search "GMO/Food and Feed Safety". For GMMs there is a specific search term "GM microorganisms". Creating a filter means that background knowledge of the domain is needed, and this knowledge must be reflected in the terms constituting the search query. To build such specialized queries expert knowledge is required to formulate ontologies. An ontology defines domain concepts and relations between them in a formal way, such that they allow automated reasoning. They are expressed in web standards such as RDFS/OWL/SKOS. Ontologies are typically organised around 'is-a' relations (taxonomy), and concepts have unique identifiers (URIs), meaning that they can be found by anyone on the web.

The following efforts can complement or upgrade the existing EFSA search strategy:

1. Improve search queries by using ontologies.
2. Select a set of sources specific for the objective set.
3. Use domain knowledge to rank search results

4.1 Use ontologies to improve (creating) search queries

Presently, a search query in EMM is crafted manually, using a set of uncoupled terms (strings) combined with logical operators. However, the process of selecting proper terms remains implicit, making it highly dependent on the experts involved. Ontologies in the area of GMO and its applications would provide a means to model the required expert knowledge explicitly, and to combine efforts of many contributors. Moreover, a more or less stable conceptual model emerges, from which specialized search queries can be extracted depending on the circumstances (for example, given a specific country to observe and its culture).

As stated above, ontologies define uniquely identifiable concepts and relations between these concepts. Each concept can have multiple labels by which it is understood by human readers. Labels can express synonyms, translations, common names (even if used wrongly), preferred names, etc. For example, the concept *GeneticallyModifiedOrganism* may have labels *GMO, GMM, GMC, GMF.* Text mining applications such as EMM can use these labels to create search queries. Secondly, the taxonomy of concepts allows automatic generation of more specific concepts. For example, 'fish' is a label of the concept #Fish. Sub-concepts of Fish are for example #Salmon, #Trout, #Herring. A search that contains the term 'fish', can be extended automatically with labels such as 'salmon', 'zalm' (Dutch for salmon), 'trout', 'herring' from all sub-concepts. These additional terms are added to the search query using the OR-operator, since any of these terms refers to fish in general.

However, using ontologies brings much more, in particular in the long run when information suppliers use them to *provide* information. We return to this in the discussion section. On the short term, using ontologies to describe the GMO domain have the following advantages.

- They help maintaining a consistent structure(s) of concepts worldwide, allowing users worldwide to add concepts and labels. Relevant GMO labs from all over the world will be asked to suggest additional concepts and labels.
- One can build on existing ontologies (which growing rapidly as a part of the Open Linked Data initiatives). The area of bioinformatics is very active in developing and applying ontologies, which is a great source for ontologies on GMO.
- Ontologies can suggest concepts (and the labels associated with them) to experts who define queries. This can be done through entry fields with auto completion or by providing ontology browsers.
- Concept labels be set in multiple languages and for each translation the language can be specified.
- Ontologies can support automated query expansion using synonyms, translations but also more specific terms as described above.

4.1.1 Improve existing keywords of GMO filters using an ontology on GMO

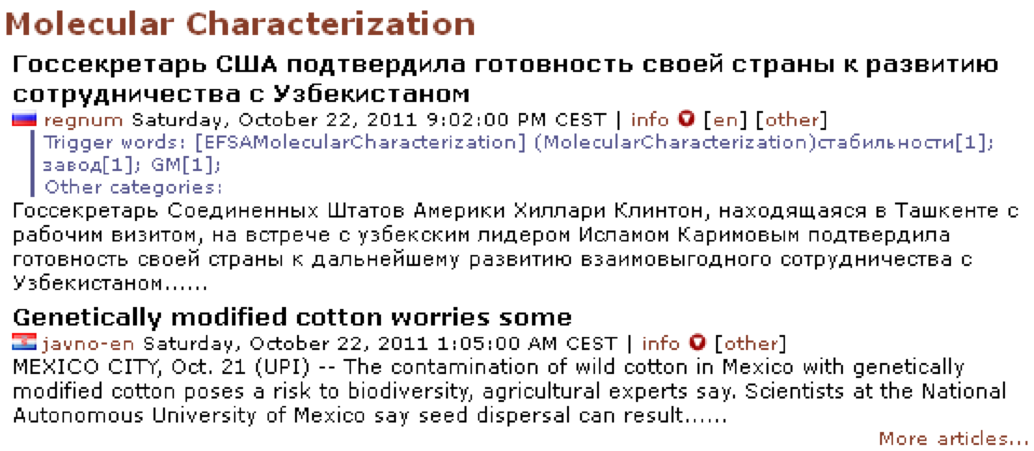
In the current EMM MedISys dedicated to EFSA, five categories have been defined (Food and feed safety (including GM plants), GM animals, GM environment, GM microorganisms, Molecular Characterization). To extend this set of filters, we suggest adding a new filter for GMOs without positive safety assessment, based on an ontology of this domain. This filter is sub-divided in GM plants, GM animals and GMM. It should be made clear to the user that it is not a search for GMOs in general, but a dedicated search for those GMOs that do not have a positive safety assessment. This means that specialized search queries are being used, requiring expert knowledge and experience.

In close collaboration with EMM (EFSA/JRC), the ontology for GMO can be set-up. We have already constructed an initial ontology using the terms described elsewhere in this report and which can be found in the Annexes. This ontology provides a platform for discussing the knowledge required for detecting new GMOs. With the knowledge of RIKILT and WUR-FBR, it is possible that combined efforts can improve the ontology. This in turn will help improving queries that are currently in use and specific queries for dedicated filters. Increase language covering

GMOs can be found world-wide. In many countries, a solid legislation is in place that regulates the approval of GMOs. Some countries or areas, however, are hot spots for emerging GMOs without any (positive) safety assessment. In order to be awareness of the existence of these GMOs, it is pivotal to provide the list of terms in these languages as well. The EMM search engine is capable of searching in 42 different languages (Table 1). The outcome is presented in the original text from the source, but translation is provided by EMM by means of Google translate (Figure 4).

Table 1. All languages available in EMM.

af - Afrikaans	et - Eesti keel	ku - Kurdî, كوردی	sl - Slovenščina
am - Amharic	eu - euskara	lb - Lëtzebuergesch	sq - Shqip
ar - العربية - Arabic	fa - فارسی - Persian	lt - Lietuvių kalba	sr - српски - Serbian
az - Azərbaycan dili	fi - Suomi	lv - Latviešu valoda	sv - Svenska
be - Беларуская - Belarusian	fr - Français	mk - македонски - Macedonian	sw - Kiswahili
bg - български - Bulgarian	he - עברית - Hebrew	mt - Malti	ta - Tamil
bs - Bosanski jezik	hr - Hrvatski	nl - Nederlands	tg - тоҷикӣ, toğikī, تاجیکی
ca - Català	hu - Magyar	no - Norsk	th - ไทย - Thai
cs - Čeština	hy - Հայերեն - Armenian	pap - Papiamento	tr - Türkçe
da - Dansk	id - Bahasa Indonesia	pl - Polski	uk - українська - Ukrainian
de - Deutsch	is - Íslenska	pt - Português	ur - اردو - Urdu
dv - Divehi	it - Italiano	ro - Română	vi - Tiếng Việt
el - Ελληνικά - Ellinika	ja - 日本語 - Japanese	ru - Русский - Russian	zh - 中文 - Chinese
en - English	ka - ქართული - Georgian	sa - Sanskrit	
es - Español		se - Swedish	
		sk - Slovenčina	


Molecular Characterization

Госсекретарь США подтвердила готовность своей страны к развитию сотрудничества с Узбекистаном

regnum Saturday, October 22, 2011 9:02:00 PM CEST | info [en] [other]

Trigger words: [EFSAMolecularCharacterization] (MolecularCharacterization)стабильности[1]; завод[1]; GM[1];

Other categories:

Госсекретарь Соединенных Штатов Америки Хиллари Клинтон, находящаяся в Ташкенте с рабочим визитом, на встрече с узбекским лидером Исламом Каримовым подтвердила готовность своей страны к дальнейшему развитию взаимовыгодного сотрудничества с Узбекистаном......

Genetically modified cotton worries some

javno-en Saturday, October 22, 2011 1:05:00 AM CEST | info [other]

MEXICO CITY, Oct. 21 (UPI) -- The contamination of wild cotton in Mexico with genetically modified cotton poses a risk to biodiversity, agricultural experts say. Scientists at the National Autonomous University of Mexico say seed dispersal can result......

More articles...

Figure 4. Example of output (October 25th, 2011) in the GMO filter. Here, the output was triggered by Russian keywords and the output is also in Russian. [en] allows a UK-English translation by Google translate.

With respect to the emergence of previous unauthorised rice GMOs like LLRice604 (unapproved LLRice601), KeFeng6, TT51-1 (hybrid Shanyou 63 and its parent Huahui No. 1), and KMD1, it is worthwhile to set a priority to the languages from the Asian region (Chinese, Japanese, Korean, Indian). Other areas of interest are the Russian Federation, the middle and South American region (Spanish, Portuguese) the USA (LLRice601, LLRice604, Bt10 maize, GM salmon), Iraq and Pakistan (GM rice) and countries in the African continent. This list of countries is not meant as a limitation for the search engine, but merely to give an indication as to what languages could be relevant to include in the multi-lingual filter. The EFSA EMRISK unit could provide support in the translation, but also relevant institutes in the respective language areas. Translation to a certain language would only be relevant when internet sites in this language will be provided.

When search terms are organised around ontologies it becomes feasible to get input for translations worldwide. GMO labs in non-English speaking countries can be asked to provide labels in their native language to increase language covering. Preferred languages would be US-English, Chinese, Russian, Japanese, Korean, Indian, Thai, Spanish, Portuguese and the obvious European languages. The added value of such a survey is the increase in multilingual concepts and the increase in country-specific internet sites that involve GMOs. In paragraph 4.6, considerations will be presented to prevent an increase in nonspecific and therefore unwanted hits.

4.1.2 Test filters for GM plants and animals to fine-tune MedISys with separate search options for official sources versus media sources

Once a filter is in place for the additional categories defined above, the filters need to be tested in a test environment to ascertain their functionality and capability to identify the correct GMOs. One option is to spike an internet page with keywords in order to see whether the EMM recognises this as an item on GMOs without a (positive) safety assessment. The validity of the terms for each filter should be tested in close collaboration with JRC and EFSA. Linge and Belyaeva, who customised MedISys for the monitoring of food and feed hazards (Linge and Belyaeva 2011) performed a training on AlertEditor and NewsDesk tools in Ispra.

4.2 Improve source selection: propose new sources related to GMOs

In the field of genetically modified organisms, in specific for plants, animals and algae, a number of sources is available that report in more or less frequency about the occurrence of GMOs worldwide. In the attached Excel file, a first attempt has been made to compile a few relevant sources. See List E (Areas and internet sites) in the attached Excel file (EMM GMO lists date.xlsx).

4.3 Improve ranking: create a procedure that removes duplicate retrievals and prioritise the outcome

Duplicate retrievals can occur due to identical output as a result of the search strategy, but also if different sources that cite the same news item. The prerequisite to remove duplicate retrievals is likely to already be included in the current EMM algorithm. Ranking of the outcome is relevant in case the number of retrievals exceeds a certain number. Only then a user can comprehend the contents and rely on the output order in such a way that the most relevant retrievals appear on top of the list. Besides ranking on relevance of the content, the EMM output is also categorised on time. For example, the items are grouped in hours/days/weeks/months ago, depending on the frequency of the news items.

4.4 Lemmatisation and stemming

One additional option to increase the full extent of list of terms (possibly derived from the labels of ontological concepts) is the use of lemmatisation (http://en.wikipedia.org/wiki/Lemmatisation) and stemming. Lemmatisation is the process of grouping together the different inflected forms of a word so they can be analysed as a single item. Inflection is the modification of a word to express different grammatical categories such as tense, grammatical mood, grammatical voice, aspect, person, number, gender and case (http://en.wikipedia.org/wiki/Inflection). Conjugation is the inflection of verbs; declension is the inflection of nouns, adjectives and pronouns. Lemmatisation is closely related to stemming. The difference is that a stemmer operates on a single word without knowledge of the context, and therefore cannot discriminate between words which have different meanings depending on part of speech. However, stemmers are typically easier to implement and run faster, and the reduced accuracy may not matter for some applications.

For instance:

- The word "walk" is the base form for word "walking", and hence this is matched in both stemming and lemmatisation.
- The word "better" has "good" as its lemma. This link is missed by stemming, as it requires a dictionary look-up.
- The word "meeting" can be either the base form of a noun or a form of a verb ("to meet") depending on the context, e.g., "in our last meeting" or "We are meeting again tomorrow". Unlike stemming, lemmatisation does select the right lemma depending on the context.

Examples taken from http://en.wikipedia.org/wiki/Lemmatisation.

In the excerpt (Figure 5) of the Web of Knowledge literature search, it is shown that lemmatization is used to find alternative forms of search terms. Here, also UK-English and US-English is automatically recognized (www.webofknowledge.com).

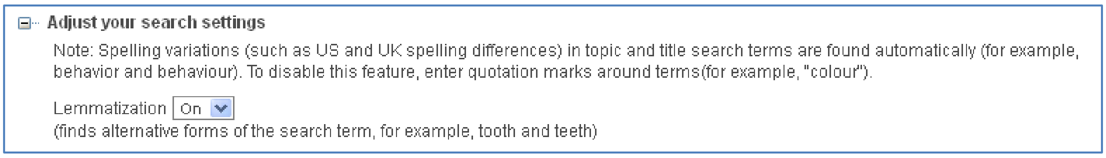


Figure 5. Lemmatization in Web of Knowledge.

We suggest that in close collaboration with EMM (including EFSA/JRC), to check the use of lemmatisation and stemming as an extension to using ontologies.

5 Discussion

From our analysis it became clear that EMM may be a useful tool to find on the WWW new reports of GMOs that are not authorised in Europe or do not have undergone a safety assessment. However, limitations are also apparent. The EMM search engine uses keywords and this has its limitation in regard to efficiency. An ontology search would, in theory, be more efficient but this approach needs to address a number of issues to prove its value.

First, the search for yet unknown GMOs is particularly hard because they can only be found indirectly. These GMOs do not have officially accepted names; publications will not state explicitly that a GMO is not approved. In this case it is very important to understand why people disclose information about such GMOs at all. Who are these people (scientists, farmers, etc.), what kind of terminology do they typically use, which countries are most active? Analysis of known sources is needed to setup more effective search queries.

Another issue is that constructing proper ontologies - which meet the needs of a specific application - is labour intensive. We have developed tools to create ontologies using sources on the web and by cooperation of multiple experts. These tools can be used to set up effective GMO ontologies.

A third issue is that presently EMM uses fixed search queries to define filters. However, an intelligent searcher adapts his strategy depending on the hits found so far. Heuristic mechanisms can be used to set up dynamic search strategies. The ontology developed for this purpose can serve as a starting point for defining heuristics.

Fourth, search engines use the statistical properties of the sources (word frequencies, distances between words, references) to rank the hits found. However, when using an ontology to generate search queries neighbouring concepts in the ontology can be used to measure to quality of a hit. For example, a document about rice is more useful if it also mentions a genetic component it contains. The ontological relation #Contains_component' between #Crop and #Genetic_component is used to give this document a higher ranking.

Finally, a weak link in the entire search process is the fact that it still depends on matching of literal strings. Using ontologies to set up search strings adds semantics at query construction time, but it is lost once the query is submitted as a 'meaningless' string of characters. A major step forward would be to have suppliers of information provide semantically rich information, referring to shared concepts available in ontologies on the web. Rather than using string matching, search queries then become logical database queries (using SPARQL, the web equivalent of SQL). This already happens on some places on the web using microformats and RDFa, embedded in HTML. Promoting semantically rich information on the web is a long term effort, but has high returns.

6 Conclusions

The Europe Media Monitor (EMM) is a powerful tool to facilitate a dedicated search for genetically modified organisms like plants, animals and micro-organisms that have no or a negative safety assessment. Important prerequisites are the careful selection of included and excluded terms/keywords. We propose to improve the current EMM GMO filters with improved set of keywords and to develop ontologies for the GMO domain to improve the search queries. In addition, selecting proper sources on the web for this specific purpose and ranking hits using the ontology will increase both recall and precision of the resulting findings. EMM is currently equipped with a tab from EFSA, containing different filters for GMOs. For the intended use as described in this report an extension of the filters seems feasible. Within RIKILT, prerequisites have been identified and initial search criteria have been suggested in order to start the discussion with JRC/EFSA for a possible improvement of the current GM-related filters, and the development of a filter for those GMOs that do not have a positive safety assessment.

References

CERA. "GM Crop Database. Center for Environmental Risk Assessment (CERA), ILSI Research Foundation, Washington D.C." Retrieved August, 2011, from http://cera-gmc.org/index.php?action=gm_crop_database.

EC. "GM-register." Retrieved August, 2011, from http://ec.europa.eu/food/dyna/gm_register/index_en.cfm.

EC. (2003). "EC Regulation No. 1829/2003." from http://eur-lex.europa.eu/pri/en/oj/dat/2003/l_268/l_26820031018en00010023.pdf.

EC. (2008). "EC Decision (2008/289/EC) on Bt63." from http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2008:096:0029:0034:EN:PDF.

EFSA (2011). Guidance document for the risk assessment of genetically modified plants and derived food and feed by the Scientific Panel on Genetically Modified Organisms (GMO) - including draft document updated in 2008 . EFSA Journal 2011; 9(5): 2150 [37 pp.]. Avaliable from: http://www.efsa.europa.eu/en/efsajournal/doc/2150.pdf.

EFSA. (2012). "Register of questions (with unit filter GMO)." from http://registerofquestions.efsa.europa.eu/roqFrontend/questionsListLoader?panel=GMO&questiontype=2.

Ewen, S.W.B. and A. Pusztai (1999). "Effect of diets containing genetically modified potatoes expressing Galanthus nivalis lectin on rat small intestine." Lancet 354(9187): 1353-1354.

Grohmann, L. and D. Made (2009). "Detection of genetically modified rice: collaborative validation study of a construct-specific real-time PCR method for detection of transgenic Bt rice." European Food Research and Technology 228(3): 497-500.

Hiwasa-Tanase, K., M. Nyarubona, et al. (2011). "High-level accumulation of recombinant miraculin protein in transgenic tomatoes expressing a synthetic miraculin gene with optimized codon usage terminated by the native miraculin terminator." Plant Cell Reports 30(1): 113-124.

ISAAA. "GM Approval Database." Retrieved August, 2011, from http://www.isaaa.org/gmapprovaldatabase/.

Linge, J.P. and J. Belyaeva (2011). Customization of MedISys for the monitoring of food and feed hazards. Service Level Agreement between European Food Safety Authority and DG Joint Research Centre (DG JRC) JRC Contract Nr. 31826-2010-07 TPW ISP Ref.: SLA/EFSA-JRC/2010/EMRISK /01, JRC: 48.

Prins, T.W., J. P. van Dijk, et al. (2008). "Optimised padlock probe ligation and microarray detection of multiple (non-authorised) GMOs in a single reaction." BMC Genomics 9.

Ruttink, T., R. Demeyer, et al. (2010a). "Molecular toolbox for the identification of unknown genetically modified organisms." Analytical and Bioanalytical Chemistry 396(6): 2073-2089.

Ruttink, T., D. Morisset, et al. (2010b). "Knowledge-technology-based discovery of unauthorized genetically modified organisms." Analytical and Bioanalytical Chemistry 396(6): 1951-1959.

Singh, C.K., A. Ojha, et al. (2008). "Detection and characterization of recombinant DNA expressing vip3A-type insecticidal gene in GMOs - standard single, multiplex and construct-specific PCR assays." Analytical and Bioanalytical Chemistry 390(1): 377-387.

Su, C.Q., J.J. Xie, et al. (2011). "Integrated structure and event-specific real-time detection of transgenic cry1Ac/SCK rice Kefeng 6." European Food Research and Technology 232(2): 351-359.

Tamas, C., B.N. Kisgyorgy, et al. (2009). "Transgenic approach to improve wheat (Triticum aestivum L.) nutritional quality." Plant Cell Reports 28(7): 1085-1094.

Wright, T.R., G.M. Shan, et al. (2010). "Robust crop resistance to broadleaf and grass herbicides provided by aryloxyalkanoate dioxygenase transgenes." Proceedings of the National Academy of Sciences of the United States of America 107(47): 20240-20245.

Yao, J.H., Y.Z. Pang, et al. (2003). "Transgenic tobacco expressing Pinellia ternata agglutinin confers enhanced resistance to aphids." Transgenic Research 12(6): 715-722.

Annex I
EU-authorised GMOs

Cotton	(281-24-236x3006-210-23)
Cotton	(GHB614)
Cotton	(LLCotton25)
Cotton	(MON1445)
Cotton	(MON15985 x MON1445)
Cotton	(MON15985)
Cotton	(MON531 x MON1445)
Cotton	(MON531)
Maize	(1507x59122)
Maize	(59122x1507xNK603)
Maize	(59122xNK603)
Maize	(Bt11)
Maize	(Bt11xGA21)
Maize	(Bt11xMIR604)
Maize	(Bt11xMIR604xGA21)
Maize	(DAS1507)
Maize	(DAS1507xNK603)
Maize	(DAS59122)
Maize	(GA21)
Maize	(MIR162)
Maize	(MIR604 xGA21)
Maize	(MIR604)
Maize	(MON810)
Maize	(MON863 x MON810)
Maize	(MON863 x NK603)
Maize	(MON863)
Maize	(MON863xMON810xNK603)
Maize	(MON88017)
Maize	(MON88017xMON810)
Maize	(MON89034 xMON88017)
Maize	(MON89034 xNK603)
Maize	(MON89034)
Maize	(NK603 x MON810)
Maize	(NK603)
Maize	(T25)
Oilseed rape	(GT73)
Oilseed rape	(T45)
Potato	(EH92-527-1)
Soybean	(A2704-12)
Soybean	(A5547-127)
Soybean	(DP356043)
Soybean	(MON 87701)
Soybean	(MON40-3-2)
Soybean	(MON87701 x MON89788)
Soybean	(MON89788)
Sugar beet	(H7-1)
Swede-rape	(MS8, RF3, MS8xRF3)

Date assessed: 2012-10-29.

http://ec.europa.eu/food/dyna/gm_register/index_en.cfm

Annex II
UGM elements used in GM plants and GM animals

Plant	Description		Animal	Description
Mir	Miraculin		L(T)F	Lactoferrin
Vip3	Vip3-related for cry-like genes		GH	Growth Hormone
CpTi	Cowpea tripsin inhibitor		AFP	Anti-Freeze Protein
Aad	Aryloxyalkanoate dioxygenase		HGH	Human Growth Hormone
1AmA1	Amaranth Albumin		BGH	Bovine Growth Hormone
mtid	Abiotic tolerance/drought resistance		CS	Coho Salmon Growth Hormone
Gna	*Galanthus nivalis* agglutinin		YP	Yellowfin Porgy Growth Hormone
Pta	*Pinellia ternata* agglutinin		CAT	Chloramphenicol Acetyl Transferase
Xa21	*Xanthomonas oryzae* pv. *oryzae* resistance		TiGH	Tilapia Growth Hormone
Cry	Crystalline genes from *Bacillus thuringiensis*		RTGH	Rainbow Trout Growth Hormone
			RGH	Rat Growth Hormone

Annex III
Species

Known GM plant species	Synonym	Latin name
Adzuki bean		*Vigna angularis*
Alfalfa	Luzerne	Medicago sativa
Arabidopsis		*Arabidopsis thaliana*
Bean		*Phaseolus vulgaris*
Broccoli		*Brassica oleracea*
Canola	Oilseed rape, rapeseed	*Brassica napus*
Canola	Oilseed rape, rapeseed	*Brassica rapa*
Cantaloupe	Melon, muskmelon, cantaloupe, netted melon, honeydew melon	*Cucumis melo*
Carnation		*Dianthus caryophyllus*
Cauliflower		*Brassica oleracea*
Chicory		*Chicorium intibus*
Chrysanthemum		*Chrysanthemum*
Cotton		*Gossypium hirsutum*
Cowpea		*Vigna unguiculata*
Creeping bent grass		*Agrostis stolonifera*
Cucumber		*Cucumis sativus*
Eggplant	Aubergine, brinjal	*Solanum melongena*
Figwort		*Scrophularia ssp*
Flax	Linseed	*Linum usitatissimum*
Lentil		*Lens culinaris*
Maize	Corn	*Zea mays L. ssp. mays*
Papaya		*Carica papaya*
Pea		*Pisum sativum*
Petunia		*Petunia hybrida*
Plum		*Prunus domestica*
Potato		*Solanum tuberosum*
Rice	Indica, japonica	*Oryza sativa*
Soybean	Soy soya, soya bean, soja	*Glycine max*
Squash		*Cucurbita pepo*
Sugar beet		*Beta vulgaris*

Known GM plant species	Synonym	Latin name
Sunflower		*Helianthus annuus*
Sweet pepper		*Capsicum ssp*
Tobacco		*Nicotiana tabacum*
Tomato		*Solanum lycopersicum*
Torenia		*Torenia spp*
Wheat		*Triticum aestivum*

Known GM animal species	Synonyms	Latin name
Mammals and birds		
Cattle	Cow	*Bos primigenius taurus, Bos primigenius indicus, Bos taurus*
Pig		*Sus domestica*
Goat		*Capra aegagrus hircus*
Sheep		*Ovis aries*
Japanese quail	coturnix quail	*Coturnix japonica*
Chicken	rooster, cock, broiler	*Gallus gallus domesticus*
Fish		
Atlantic salmon		*Salmo salar*
Coho salmon		*Oncorhynchus kisutch*
Chinook salmon		*Oncorhynchus tshawytscha*
Tilapia		*Oreochromis spp.*
Medaka		*Oryzias latipes*
Zebra fish		*Brachydanio rerio*
Common carp		*Cyprinus carpio*
Channel catfish		*Ictalurus punctatus*
African catfish		*Clarias gariepinus*
Rainbow trout		*Oncorhyncus mykiss*
Cutthroat trout		*Oncorhyncus clarki*
Goldfish		*Carrassius auratus*
Northern pike		*Esox lucius*
Loach		*Misgurnus anguillicaudatus*
Sea bream		*Sparus aurata*
Red Sea Bream		*Pagrus major*
Blunt snout bream		*Megalobrama amblycephala*

Known GM animal species	Synonyms	Latin name
Nigorobuna		*Carrassius auratus grandoculis*
Walleye		*Stizostedion vitreum*
Others		
Brine shrimp		*Artemia franciscana*
Seaweed		*Laminaria japonica*
Seaweed		*Undaria pinnatifida*
Sea Urchin		*Strongylocentrotus purpuratus*
Sea Urchin		*Paracentrotus lividus*
Sea Urchin		*Arbacia lixula*
Abalone		*Haliotus rufescens*

Annex IV
Internet sites

Description	Site	Language	Remarks
Scientific literature			
Web of Knowledge, Web of Science	http://apps.webofknowledge.com		Thomson Reuters
Google Scholar	http://scholar.google.nl		Literature search engine from Google
Government			
Australian Government Department of Health and Ageing	http://www.ogtr.gov.au/	UK	Office of the Gene Technology Regulator Website
Canadian Food inspection agency	http://www.inspection.gc.ca/english/toce.shtml		
Biosafety Clearing House (BCH)	http://bch.cbd.int/	UK	
Indian GMO Research Information System (IGMORIS)	http://igmoris.nic.in/	UK	
CTN Bio	http://www.ctnbio.gov.br/	PT, UK	
GMO info China	http://gmo.apqchina.org/cn/product/list.asp	CN	
SE-CIBIOGEM Mexico	http://www.cibiogem.gob.mx	ES, UK	http://sippic.main.conacyt.mx:7777/pls/sippic/rnbioseg.sol_liberacion
BCH China	http://english.biosafety.gov.cn/	UK	
Health Canada	/http://www.hc-sc.gc.ca/fn-an/gmf-agm/index-eng.php	UK, FR	Genetically Modified (GM) Foods and Other Novel Foods
Food standards Australia and New Zealand	http://www.foodstandards./gov.au	UK	
Belgian Biosafety Server	http://www.biosafety.be/	UK	
Genetic Engineering Approval Committee (GEAC), Ministry of Environment and Forests, India	http://moef.nic.in/modules/project-clearances/geac-clearances/	UK	
Science net China	http://www.sciencenet.cn/english/	CN, UK	
Rapid Alert System for Food and Feed (RASFF)	https://webgate.ec.europa.eu/rasff-window/portal/	UK	
Belarus Telegraph Agency	http://news.belta.by/en	RU, UK	
National Institute of Health Sciences	http://www.nihs.go.jp	JP, UK	

Description	Site	Language	Remarks
NGOs			
Gentech.NL	http://www.gentech.nl/	NL	
Grain	http://www.grain.org/	UK, ES, FR	
Greenpeace	http://www.greenpeace.org/international/en/	UK	
GM contamination register	http://www.gmcontaminationregister.org	UK	
GM Freeze	http://www.gmfreeze.org/	UK	the campaign on GM food, crops and patents
Genewatch UK	http://www.genewatch.org/	UK	
Excellence Through Stewardship	http://www.excellencethroughstewardship.org/	UK	Advancing best practices in agricultural biotechnology
Save our seeds	http://www.saveourseeds.org/	DE, UK	
AgBioWorld	http://www.agbioworld.org/	UK	A daily collection of news and commentaries on agricultural biotechnology
GMO Compass	http://www.gmo-compass.org/eng/home/	UK	
Information Systems for Biotechnology (ISB)	http://www.isb.vt.edu/	UK	a national resource in Agbiotech information
African Centre for Biosafety (ACB)	http://www.biosafetyafrica.org.za/	UK	
GM Watch	http://www.gmwatch.org	UK	
Say no to GMOs	http://www.saynotogmos.org	UK	
GMO free regions	http://www.gmo-free-regions.org	UK	
Internet in general			
ISAAA	http://www.isaaa.org/	UK	multilingual by google translate
SeedQuest	http://www.seedquest.com/	UK	Global Information Services for Seed Professionals
Delat Farm Press	http://deltafarmpress.com/	UK	Timely reliable information for delta agriculture
ArborGen	http://www.arborgen.us	UK	Tree improvement
Biosafety information centre	http://www.biosafety-info.net/	UK	
Environmental news network	http://www.enn.com/	UK	
The Parliament.com	http://www.theparliament.com	UK	
European Public Health Alliance	http://www.epha.org/	UK	
China Daily	http://ipr.chinadaily.com.cn		China Intellectual property
Euginius GMO database	http://www.Euginius.com		

Description	Site	Language	Remarks
CERA GM crop database	http://www.cera-gmc.org		ILSI Research Foundation
GMO detection method database	http://gmdd.shgmo.org		GMO Detection Laboratory in Shanghai Jiao Tong University (GMODL-SJTU)
The Examiner	http://www.examiner.com	UK	
Rice news	http://oryza.com/Rice-News/	UK	
EurekAlert China	http://chinese.eurekalert.org	CN, UK	
Next Big Future	http://nextbigfuture.com	UK	
Agropages	http://news.agropages.com	CN, UK, IN	World Agrochemical pages online
Before it's news	http://beforeitsnews.com	UK	
	http://wenku.baidu.com	CN	
Vatgia Vietnam news	http://www.vatgia.com/home/		
Korea	http://media.daum.net		
Mongolia	http://tsagtur.mn		
Lietuvių kalba	http://www.straipsniai.lt		
MD Today, Korea	http://www.mdtoday.co.kr		
Feedage.com	http://www.feedage.com	UK	
GeneRef	http://www.generef.com/	UK	Health and Science News, Genomics, Stem Cells, Nanotechnology: A Global Resource
The Voice of Russia	http://english.ruvr.ru/	UK, RU	The Voice of Russia radio website
Rice science	http://www.ricescience.org	UK, CN	
Biotechnology headlines	http://www.biotechnology headlines.com/	UK	
GM database	http://gmdb.nihs.go.jp	JP	Recombinant GM database on animals, GMM, industrial raw materials, medicine
de Volkskrant	www.volkskrant.nl	NL	Dutch national newspaper
Elsevier	www.elsevier.nl	NL	Dutch national newspaper
Gezondheidskrant	http://www.gezondheids krant.nl	NL	Dutch topic newspaper
Nederlands Dagblad	http://www.nd.nl	NL	Dutch national newspaper
RTV Rijnmond	http://www.rijnmond.nl	NL	Dutch broadcast news site
Telegraaf	http://www.telegraaf.nl	NL	Dutch national newspaper
Nieuwsbank	http://www.nieuwsbank.nl	NL	Dutch national news site
Algemeen Dagblad	http://www.ad.nl	NL	Dutch national newspaper
Food safety info Japan	http://www.fsafety-info.org	JP	
ZHTC	http://new.zhtech.cn	CN	

Annex V
Generic keywords

Keywords to be included

Adventitious+presence
AFP
Agrobacterium
Antifreeze+protein (afp)
Aqua bounty
Aquadvantage
Assessment
Basf
Basmati
Bayer
Biolistic
Bio-protein+production
Biotec
Biotechnological
Biotechnology
Carnation
Cassette
Cattle
Cisgen%
Cisgenesis
Cloning
Construct
Creeping+bent+grass
Crop+improvement
Cultivar%
Dedicated+biosensors
Drought
Engineered
Enhanced
Event
Gene
Genetically+improved
Genetically+modified
Glowfish
Gm
GMC (=GM crop)
GMF (=GM fish)
GMM
GMO
Goat
Hazard+characterization
Improved+livestock+production
Intragenic
Low+level+presence
Marker+free
Marker+gene
Molecular
Monsanto
Novel
Oecd uid
Ornamental+fish
PCR
Petunia
Pig
Plasmid
Rainbow+trout
Resistance
Resistant
Risk+assessment
Risk+characterization
Roundupready
Safety
Salinity
Salmon
Salmon+growth+hormone (gh)
Sentinel+fish
Shotgun+cloning
Stack
Stacked
Stacking
Substantial+equivalence
Targeted
Tolerance
Tolerant
Transformation
Transformed
Transgenic
Transgenic+animal%
Transgenic+crop%
Transgenic+microorganism%
Trout
Unauthorised
Unauthorized
Vector

Keywords to be excluded

Galactic+Money+Organizer
Gastric+Medical+Operation
General+Manager
General+Medical+Officer
General+Moly+Inc
General+Motor%
Girl+Make+Out
Gitaar+en+Mandoline+Orkest
GMO+internet+group (=www.gmo.jp)
GMO+LLC
GMO+mutual+funds
GMO+webmail
GMO+free
Government+Manure+Office
Graduate+Marketing+Organization
Grantham, Mayo, Van Otterloo & Co. LLC
Grants Management Officer
Greater Milwaukee Open
Green+Media+Organization
Group+Management+Office
Guest+Memory+Operation
Gum+Made+Ordinarily
Illinois+Central+Railroad+Company
Non-GM
Non-GMO
XA-GMO
Buick
Cadillac
Cars
Chevrolet
Detroit
Farm
Fishing
Opticasgmo (=http://www.opticasgmo.com)
Webfeed%

RIKILT Wageningen UR is part of the international knowledge organisation Wageningen University & Research centre. RIKILT conducts independent research into the safety and quality of food. The institute is specialised in detecting and identifying substances in food and animal feed and determining the functionality and effect of those substances.

RIKILT advises national and international governments on establishing standards and methods of analysis. RIKILT is available 24 hours a day and seven days a week in cases of incidents and food crises.

The research institute in Wageningen is the National Reference Laboratory (NRL) for milk, genetically modified organisms, and nearly all chemical substances, and is also the European Union Reference Laboratory (EU-RL) for substances with hormonal effects.

RIKILT is a member of various national and international expertise centres and networks. Most of our work is commissioned by the Dutch Ministry of Economic Affairs and the Netherlands Food and Consumer Product Safety Authority. Other parties commissioning our work include the European Union, the European Food Safety Authority (EFSA), foreign governments, social organisations, and businesses.

More information: www.wageningenUR.nl/en/rikilt