A Model-based Information Security Risk Assessment Method for Science Gateways

Evert Mouw
University of Amsterdam
The Netherlands
Email: post@evert.net

Guido van 't Noordende
University of Amsterdam
The Netherlands
Email: guido@science.uva.nl

Baas Louter
University of Amsterdam
The Netherlands
Email: baas.louter@amc.uva.nl

Silvia Delgado Olabarriaga
University of Amsterdam
The Netherlands
Email: s.d.olabarriaga@amc.uva.nl

***Abstract*—BACKGROUND: Information Security is important for e-Science research groups and other small organisations that design and operate science gateways and virtual research environments, especially when such environments are being used for (bio)medical research. We propose a novel method to do risk assessments: MISRAM, the Model-based Information Security Risk Assessment Method. It uses an information architecture model, a method to assign values to information assets and IT components, and a method to calculate risks. The output of MISRAM is a ranked list of risks and a list of actionable tasks to solve the main issues.**

METHODS: MISRAM was applied as a test case to an e-Science research group at a Dutch research hospital. Meetings and surveys were used to create and evaluate lists of information assets and IT components. One meeting was used to create a list of practical task recommendations.

RESULTS: Good insight into the information architecture and security problems of the IT infrastructure was gained. Also the participating group members confirmed that the identified security issues were realistic.

CONCLUSIONS: Our approach raises awareness about security among the developers and operators of e-Science environments. It also gives insight in how the technical architecture affects information security. Traditional questionnaires are an important part of any risk assessment, and MISRAM's inclusion of such generic questionnaires is an important aspect to create an integrated information security risk assessment.

Keywords*—*MISRAM, risk assessment, information security, e-Science, science gateway, DCRA, IT&T.

I. Introduction

Science gateways are community tools, typically web portals, that enable and facilitate access to distributed infrastructures such as computational Grids. These portals are developed by teams of e-Science experts in various fields (distributed computing, user interfaces, data management, visualisation, etc), and are typically operated by (subsets of) the same teams as a service for the end-users, which are experts in other scientific domains, not necessarily in computing. Typical services include high-throughput or -performance computing, visualisation, and access to data or application repositories. Science gateways are therefore complex systems that combine a large number and variety of software components under the same 'container'. Very often some of these components are packages developed by other parties or services provided remotely. For a detailed discussion about the types, technologies, and required properties of science gateways, please refer to the Science Gateway Primer recently published by the EGI-Inspire Virtual Team on Science Gateways[1].

The e-BioScience group in the Academic Medical Center (AMC) in Amsterdam, the Netherlands, offers a science gateway, coined e-BioInfra gateway, for experimental data analysis in the fields of high-resolution medical imaging, genomics, and proteomics [1]. The backend is based on scientific workflow management technology [2], and makes use of grid-based computing resources. Using such distributed resources for sensitive medical data requires a solid security policy. This is particularly true when intrinsically identifyable DNA data is involved [3]. Ideally, the gateway should be subject to Risk Management [4] following recommendations given by standards such as ISO 27002 and the related Dutch standard for information security management in healthcare, NEN 7510. An important element of both standards is the execution of a good Risk Assessment (RA), where a systematic analysis of security aspects of the system is carried out.

Concerned with the security features of the e-BioInfra gateway, and driven by the increasing awareness of biomedical researchers about privacy and ethical issues related to the information that they process through this system, it was decided to conduct a risk assessment of the e-BioInfra gateway. In this paper, our risk assessment method is presented, together with an example based on a simplified version of the risk calculations to illustrate the approach. The experience obtained by applying a modified risk assessment method at the AMC is also presented. The method worked well for this e-Science environment, providing insights and raising awareness about the security of the e-BioInfra gateway. The method could also be useful to other groups that develop and operate science gateway services.

A. Some specific needs for science gateways

The distributed nature of the data processing, the amount of data and computing, and the organisational characteristics of a science gateway are unique. A particular aspect for science gateways is that components can be internal or external information processing systems which are owned by different organisations. Therefore, the method applied should facilitate the analysis of these interdependencies, following a model-based approach.

[1] https://documents.egi.eu/public/RetrieveFile?docid=1463&version=10&filename=Science_Gateway_Primer_v092_nComments.pdf)

CEUR-WS.org/Vol-993/paper15.pdf

A typical e-Science group is small and consists of highly skilled experts who develop and maintain their own software. These properties make it hard for external security experts to carry out a Risk Assessment. In fact, such outside analysts often lack inside information and good working knowledge on the technical components used by a science gateway. Our method takes a multiple-stakeholder approach, also favoured by Zambon et al. [5], where people who manage the system and the users of the system are involved in the risk assessment.

B. Introduction to Risk Assessments

The goal of doing a Risk Assessment (RA) is to get an overview of the most relevant risks threatening some system, person or organisation. For example, when playing chess, a simple risk assessment would include the dangers associated with losing the center pawns early in the game. Risks and dangers are measured as probabilities. Even in a deterministic game like chess, risk assessments come down to probabilities [6]. Calculating such probabilities, and thus risk, can become a messy business if the situation becomes complicated [6]: "A complicated chess position requires deep calculations and is more likely to cause a human player to make an error."

A risk assessment is a method to estimate risks. In this paper, we will focus on estimating the most important risks for information systems used in science, such as science gateways. Which information security risks threaten such gateways and how to estimate them? How does one identify the main risks and causes of such risk?

Risk exists when something of value (e.g. an asset) could be lost or damaged. Many different definitions of risk are being used[2]. ISO 31000:2009 defines risk as *effect of uncertainty on objectives*. A common idea is that risk is a probability or potential that some unwanted event will happen in the future. We use the following definition: *risk is the chance of some bad event happening multiplied with the impact of that event*. So, risk is likelihood multiplied with impact costs (damage).

The "something of value", which we will call an *asset*, must be defined for the risk assessment. Assets include information (such as patient data, customer lists, software, source code, process data) and IT assets (such as server hardware). The value of an asset is linked to the impact when such an asset would become unavailable (e.g. due to a computer virus or hardware failure) or would be compromised (such as theft of data). For an information security risk assessment, we will focus solely on information assets.

Information assets are abstract assets that are stored in some concrete form. An example is the knowledge a worker has about the production process. This knowledge is an information asset, but it cannot be easily backed up or copied. Another information asset could be instructions (information) contained in a manual, which would be far easier to copy.

According to the UK National Archives [7], there are two dimensions to value an information asset:

- **Business value** for an organisation or an individual (e.g. a patient). This also includes scientific value. In this text, this will just be called *value*.
- **Confidentiality** of the information. Who should have access (scope or domain) and who should have not?

For example, the fire brigade needs to know where a fire started in order to send the fire engine to the right location. That is of high business value, so such information should be reliable and available to the fire fighters. But it isn't confidential at all. Banks holding the money of their clients do store information with both business value and confidentiality. But even if there is little business value, some information can still be confidential, e.g. bank transfers to a small savings account.

C. Vulnerability to unintentional errors and intentional attacks

An information asset can be stolen and sold (e.g. creditcard information), but it can also be damaged. A component in an IT infrastructure can be vulnerable to *intentional attacks* and also to *unintentional errors*, also called *unintentional mistakes*. Experts can indicate the confidence they have in the security of components by estimating the probability of intentional and unintentional failures. These factors are used to measure vulnerability by NIST SP 800-33 [8, p. 18] and Cisco [9]. The *total vulnerability* is a summation of the vulnerability to unintentional errors and the vulnerability to intentional attacks.

D. Model-based vs. checklist-based Risk Assessments

There are many methods to choose from to perform a Risk Assessment (RA). Most of them consist of a list of 'items' to check or score. According to Morali et al. [10], current mainstream risk assessment methods "do not take the IT architecture of the system under examination". In contrast to the checklist-based RA methods, a model-based or architecture-based method provides richer possibilities for the analysis of complex systems.

Architecture-based RA methods start from a model of an information system, taking into account aspects such as dependencies between components and value of the information manipulated by them. This makes it possible to take architectural aspects into account.

We found only four RA methods recently described in the literature which use a model-based approach, namely: CORAS [11], Qualitative Time Dependency (QualTD) [5], RiskM ('Risk Method') [12], and Distributed Confidentiality Risk Assessment (DCRA) [13]. A thorough analysis of model-based RA methods is described by Mouw [14]. We concluded on on the basis of that work [14] that the DCRA [13] is the most adequate for our use-case.

E. Enterprise layers

DCRA uses a modelling of the 'enterprise' that resembles the ArchiMate modelling language [15, p. 3] and distinguishes three main layers [13]:

[2]For more definitions of risk, see http://en.wikipedia.org/wiki/Risk

- The **Information layer** consists of **Information Assets**. [3]
- The **Application layer** supports the information layer with services that are realised by (software) applications.
- The **Technology layer** offers infrastructural services (e.g., processing, storage and communication services) needed to run applications, and is realised by computer and communication hardware and system software.

F. DCRA's IT&I model

At the heart of DCRA lies the IT assets and Information (IT&I) model. The information assets (*I*) are distributed over supporting IT assets (*IT*). For an example, see fig. 1.

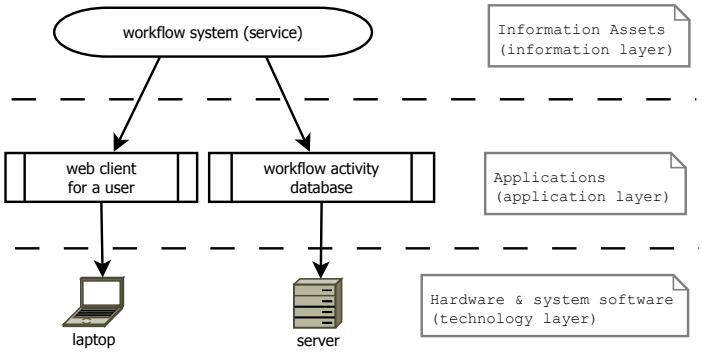

Fig. 1: Simplified example of an DCRA IT&I information architecture.

In the example (fig. 1), 100% of the information asset "workflows" could be assigned to the application "workflow activity database", and 10% of "workflows" could be assigned to the application "web client". Both of these applications, which reside in the application layer, are mapped to components in the technology layer. E.g., the web client could be running on a laptop. For DCRA, if the database server is vulnerable, then the impact is higher then when the laptop is vulnerable, because more of the information asset is mapped to the workflow database.

II. A NOVEL RA METHOD: MISRAM

Although DCRA [13] provides an excellent foundation, it is not optimised for the context of science gateways. The method is labour-intensive due to its detailed information model and somewhat complex calculations, and only measures risk of confidentiality breaches. MISRAM simplifies the calculation model for pragmatic reasons, which comes at the cost of precision reduction (see the Discussion). We extended the model to measure both the confidentiality and business value of information [7]. Furthermore, we decided to use the expertise, including implicit knowledge, of the e-Science group members to assess the security of the IT components of the science gateway. We used (and recommend) surveys to assess all components of the IT infrastructure.

Our method was *not* designed to easily find so-called "low-hanging fruit" (easy to fix problems), but to find those problems that are most likely to cause real damage. The method generates a prioritised list of issues that need to be addressed for security improvement.

We call our modified method 'MISRAM', an acronym for "Model-based Information Security Risk Assessment Method". It is based on an enterprise information model proposed by DCRA [16], and an overview of the whole MISRAM procedure is presented shortly.

To do a risk assessment for an organisation, one first needs to create an information model. The model should contain indicative numbers for business values, confidentiality indicators, attack vulnerabilities, and chances of unintentional errors. One could use surveys or other methods to obtain those numbers. And then, after the calculations are done and a ranking of relative risks is made, one needs to translate these relative risks to a risk classification, such that this information can serve as input for the risk management coordinator.

A. MISRAM: the whole picture

We have broken down the whole MISRAM process in ten steps. Part of these can be executed in parallel. See fig. 2 for a graphical representation of the workflow.

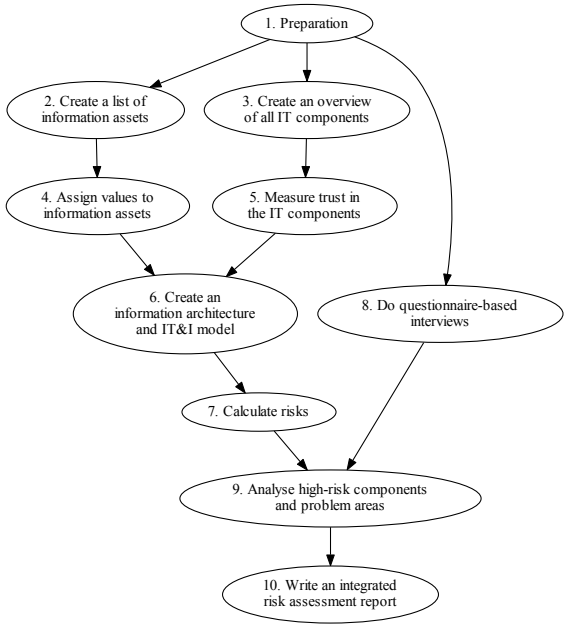

Fig. 2: Workflow of an architecture-based information security risk assessment project.

1) Preparation: Form a project team, agree on roles and responsibilities, and make a project plan. Determine the scope: what is the unit or group that will be assessed, who are the

[3] In AchiMate, the Information layer is called the Business layer. It offers products and services to external customers, which are realised in the organisation by business processes performed by business actors. In DCRA and MISRAM, this business layer is replaced by an information layer. An information layer has not all the entities a business layer has, and some are simplified. For example, actors will become abstract users.

stakeholders internal and external to that unit, and what is the environment of the unit?

2) Create a list of information assets: It is important to obtain a detailed list. The best way to collect such a list is to interview different stakeholders such as internal developers and administrators, business managers, clients, and suppliers. The detailed list should be reduced to a list that is clear, short, and has no overlaps. These information assets will comprise the information layer.

3) Create an overview of all IT components: The IT components of an organisation typically comprise (business) applications, operating (system) software, and hardware. The applications will comprise the application layer. The system software and the hardware will comprise the technology layer. Also external computation and storage facilities such as cloud or grid resources must be listed.

4) Assign values to information assets: All stakeholders should be invited to score the value of the information assets, e.g. by using a small survey. For each information asset, participants are asked to score the importance of two dimensions: *business value* and *confidentiality*. Many scoring systems are possible; we suggest a relative value on a five-point Likert scale [17], using the values *none, low, moderate, high, very high*. Also provide a *no answer* option. Additionally, you should include a question on the type of stakeholder, such as end-user or scientist or developer or manager, if you want to differentiate on stakeholders later in the process. This differentiation, however, will not be further discussed here. Finally, when using a survey, keep it short to promote participation and to avoid selection bias – busy people don't have time for extensive surveys. Always include a comment field to allow free-text comments.

The calculation of values from scores depends largely on the number of interviews. Using the median is not such a good idea because a survey often uses discrete responses. When e.g. six of the respondents assign value '3', and the seven others assign '4', then the mean (3.5) represents the average opinion better than the median (4). A 'no answer' value does not influence the mean, but many 'no answer' scores on an component could raise a red flag: the respondents don't feel sure about such a component.

In a small group, just one person with a somewhat extreme opinion can have a large influence on the mean score, which might be undesirable. A good method to reduce the influence of such outliers is to use a truncated (trimmed) mean. The $x\%$ trimmed mean leaves out the highest and the lowest values. Often a value of 20% for x is chosen, in which case the 10% highest values and the 10% lowest values are removed and the mean is taken over the remaining values[4].

Some group members might have more knowledge and experience than others. To take into account their expertise level, a question can be included in the survey asking the respondent to estimate her/his level of expertise or experience. There are various methods to classify expert levels. One well-known way to do this is the Dreyfus model of skill acquisition [18]. A modification to that model was made by Schempp, who renamed "Advanced Beginner" to "Capable" [19]. The five levels of expertise of Dreyfus-Schempp are: *Novice, Capable, Competent, Proficient, Expert*.

5) Measure trust in the IT components: Confidence or trust in the reliability or security of components that are in the application and the technology layers of the architecture can be measured by surveying or interviewing experts: the developers, administrators and expert users of those components. The same comments on designing surveys from step 4 are equally important here.

For all components, ask the experts how they assess security and reliability of a component, based on two factors: the vulnerability of the components for *intentional attacks* and the vulnerability for *unintentional errors*.

6) Create an information architecture and IT&I model: The method described in section I-F can be applied here. All main information assets and IT components should be included if their risks and effects are of interest, but it is not advisable to add too much detail; it is considered most productive to focus on the big picture. If architectural documents already exist, be careful to examine whether they are up-to-date.

7) Calculate risks: The likelihood of trouble happening is probed by the measured trust in the IT components. The costs are determined by the value and confidentiality of the information assets. The method used to calculate risks in MISRAM is described in section II-C.

The result of this process is a ranked list of information assets that are under risk (in our model, the risk of losing business value or losing confidentiality).

The calculated risk values lack a risk classification to determine whether a particular asset is in fact at 'high', 'medium', or 'low' risk. For example, if the complete system is already highly trusted by the interviewees, even the largest value coming out of the calculation might be in practice considered at "low risk". The risk classification should be carried out in e.g. a group meeting.

8) Do questionnaire-based interviews: Some security related issues fall outside the scope of the architecture-based approach. Examples are password and clean desk policies, employee screening, and so on. Furthermore, some specific issues might be required to assess because of sector-specific regulations, such as the NEN 7510 in the Dutch healthcare system. A questionnaire based on local needs and legal requirements should be created, and then used for interviews with employees.

The interviews could reveal vulnerabilities that were not found using the earlier steps. One example is the handling of user data; although the group members could in theory give a low score to the business value and/or the confidentiality of user data, often a legal liability exists.

9) Analyse high-risk IT components and problem areas: IT components cause high risk because, for example, they are not trusted or they store or process valuable or confidential information. Such high-risk components are identified in the earlier steps and they should now be further investigated. Also important vulnerabilities that emerged from the questionnaire-based interviews should likewise be given extra attention.

[4] A short introduction to trimmed means: http://en.wikipedia.org/wiki/Truncated_mean

The goal in this step is to create a list of specific, concrete security problems that can be contained or fixed by specified actions. A good way to discuss and investigate the high-risk IT components and vulnerabilities is to organise a group meeting.

10) Write an integrated risk assessment report: A risk assessment report should be written to document the methods and results and to serve as a starting point for actions directed to mitigate risk. The report should include an overview of the organisation that was assessed and the general approach that was being used, present the results for each step, and summarise the main security risks and their causes. This report should also contain a ranked list of information assets that are under risk.

B. Calculations

MISRAM first limits its goal to finding a *relative ranking* of risks. We estimate risk using the value of an information asset multiplied with the vulnerability of the IT components concerning two factors, *intentional attacks* and *unintentional errors*. A *weighted average* can be used to give more emphasis to one or the other. For example, a reason to give more emphasis to unintentional errors could be that one expects that the group members are better at scoring the risk of unintentional errors than scoring the risk of intentional attacks.

The vulnerability score Θ is calculated as follows:

$$\Theta = 2 \times ((\alpha) \times attacks + (1 - \alpha) \times errors)$$

with $\alpha \in [0.0, 1.0]$ representing the weight given to the risk of attacks relative to errors estimated by the experts. With $\alpha = 0.5$, the formula simply becomes:

$$\Theta = attacks + errors$$

C. Example of a calculation in MISRAM

We will use a simple example to clarify the method of calculation. See fig. 3 to understand the following example. We are going to calculate the risk of losing business value $risk\{v\}$. Note that $risk\{v\}$ must be calculated separately from the risk of losing confidentiality $risk\{c\}$. The latter will not be calculated in this example.

Each component has a *vulnerability score* Θ that consists of a vulnerability to *intentional attacks* and a chance of *unintended errors*. For the *server*, the equation would be:

$$\Theta_{server} = attacks + errors = 3 + 1 = 4$$

The information asset *workflows* has business value *2*. It is stored and processed using a *web client* (vulnerability score for intentional attacks = 1 and for unintentional errors = 2) with $\Theta_{webclient} = 1 + 2$, on a *laptop* with $\Theta_{laptop} = 4 + 4$. Also, all workflows are processed by a *database server* with $\Theta_{db} = 2 + 1$, which runs on the physical *server* with Θ_{server} (see above).

The total risk for the business value $risk\{v\}$ of workflows is calculated as the value of workflows (as estimated with a survey), multiplied with the summed vulnerability scores of the IT components that process or store workflows:

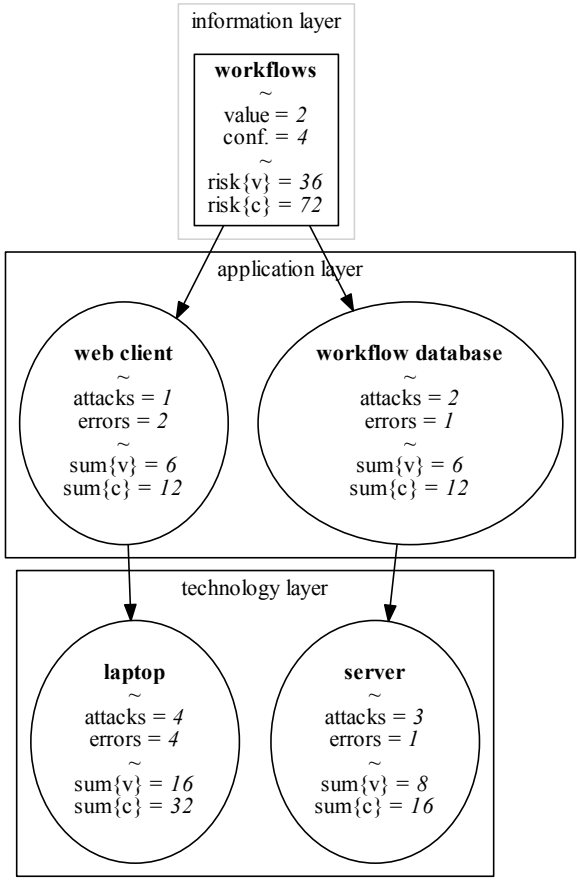

Fig. 3: Example of an DCRA IT&I risk calculation in our modified and simplified fashion. Three layers can be recognised: the boxes on top represent information assets (information layer), the ellipses in the middle represent software applications (application layer), and the bottom ellipses represent system-level software and hardware (technology layer). 'Value' is business value, 'conf.' is confidentiality, 'risk{v}' is the risk of losing business **v**alue for that information asset, and 'risk{c}' is the risk of losing **c**onfidentiality. See the main text on how the calculations are done.

$$\begin{aligned} & risk\{v\}_{workflows} \\ & = value\ of\ workflows \cdot sum\ of\ vulnerabilities \\ & = 2 \cdot\ (\Theta_{webclient} + \Theta_{laptop} + \Theta_{db} + \Theta_{server}\ ) \\ & = 2 \cdot\ ((1+2) + (4+4) + (2+1) + (3+1)) \\ & = 36 \end{aligned}$$

Doing this for all information assets will create a list of all business value risks and will show which information assets are

most at risk. The same method can be used for confidentiality risk (not shown here).

To calculate how much risk for business value a component causes, one needs to sum all the partial risks that it causes for all the information assets that it stores or processes (is connected to). In the example, *server* has a $sum\{v\}$ = *8*, because it is connected to one information asset *workflows* with value *2* and *server* itself has a vulnerability of 3+1 = *4*.

III. Case study: MISRAM applied to the e-BioInfra gateway

We tested MISRAM at the e-BioScience group at the AMC, from now on simply referred to as 'group'. With the cooperation of the group members, who were regarded as experts and were being surveyed, we assessed the information security risks for the whole infrastructure for which the group is responsible, including a science gateway. The first author of this paper was responsible for carrying out the assessment and also participated in the group as an IT administrator at the time. The group members participated in the various meetings, interviews and surveys described below. The full report on this case study [20] is not public for security reasons, but details might be requested by contacting the authors.

A. Creating an inventory and a model

A meeting was held to determine the enterprise architecture, including business services. The proposed enterprise architecture model was later modified based on individual interviews with developers. Finally, agreement was reached on the main IT components and connections between them.

The technology layer of the e-BioInfra is split in a 'physical and system' layers and an external 'grid' layer. Grid resources include computing elements and storage elements. The whole model appears as having four layers, with an additional separate grid layer, which better represents the separation of the grid services from the technical and organisational aspects of the e-BioInfra.

Another group meeting was held to identify the information assets, during which the concept of information assets was explained and discussed. Later, in private interviews, suggestions were made by some group members to add information assets or to add more detail by splitting up some assets. One group member commented on the generic nature of one information asset, arguing for more detail. Such discussions should be closed before the information asset survey starts. Otherwise, the list of information assets might change after the survey, and the survey needs to be repeated. A meeting or a group manager decision can help to determine a final list.

The enterprise architecture (modelled using ArchiMate) was combined with the information assets to create an information architecture, modelled using IT&I. Most group members understood the IT&I without much explanation. The IT&I diagram was created using Archi (http://archi.cetis.ac.uk/).

Then, a survey was carried out to assign business values and confidentiality values to the information assets. Most group members (5) participated in the survey.

Another survey asked the group members to give their opinion on the risk or vulnerability of (unintentional) errors and (intentional) attacks associated with the IT components in the application and technology layers.

B. Identifying risks

A simple C++ program was developed to calculate the risks from the IT&I relations, the survey data about the business value and confidentiality of the information assets and the vulnerability to errors and attacks of the IT components. Trimmed means were used to compensate for outliers. From the ranked list of calculated risks, it became clear that three IT components have the highest cumulative risk:

1) The server that hosts the science gateway software and other e-BioInfra applications, including a workflow engine.
2) The external computations done on the grid.
3) The workflow system itself.

The server itself scored very reasonably on the vulnerability survey, as the group experts showed trust in its security. Still, the server ended up as one of the main causes of risk because of all the applications it hosts.

A generic questionnaire based on NEN 7510 was also applied. It helped to identify a number of additional security issues, mostly related to behaviour and to organisational policies, such as account management and desktop policies. A few major problems became clear from the generic questionnaire, which could be solved by these concrete actions:

- All group members should lock their screens when they leave their desk. Controls and incentives should be put in place to enforce this.
- All mobile devices should use encrypted storage.
- Let those not on payroll and apprentices sign an non-disclosure agreement. Note that this is standard procedure for AMC employees.
- When an employee leaves, accounts should be promptly revoked.
- There should be a reporting and logging facility for security incidents.
- Web applications should use SSL (HTTP*S*).

Although these actions seem quite obvious, they sometimes are not carried out in academic settings, where mutual trust and intellectual curiosity characterise the culture on the workplace.

The results of the risk calculation and the questionnaire were discussed during a meeting. The group members agreed that the three identified high-risk IT components were important, and voiced their trust in the method used to identify the components and formalise the implicit knowledge in the group. A few actionable measures were proposed, but also a small generic discussion on information security started. The meeting helped to raise awareness and to identify the most important practical tasks.

IV. Discussion

The proposed MISRAM method has been successfully applied at the e-Science group at the AMC to assess the risks of the e-BioInfra platform and its science gateway. The group was satisfied with the insights delivered by MISRAM, which indicated practical measures to better address the security regulations for digital processing of medical data. Although we are encouraged by this first result, MISRAM needs to be tested with other groups and other types of scientific gateways for a more conclusive evaluation. Below we present a reflection of aspects that influence the method's performance.

A. Usage of expertise in e-Science groups

MISRAM makes use of the (implicit) knowledge of the people that develop, administer and support the e-BioInfra science gateway. Information assets and component vulnerabilies can be determined by interviewing or surveying these group members.

Typical *e-Science* groups do their own development and systems administration, and employ a team of experts on the IT components of their own architecture. These experts build the (software) infrastructure that stores and processes the information. Such information includes both assets of the group itself and information assets from external parties (users).

The expertise of the developers of a science gateway on the local security situation might exceed the expertise of an external security expert. Their knowledge on errors within the technical infrastructure is based on their day-to-day experience, so the infrastructure experts can be trusted to have a good feeling for the vulnerability to *unintentional errors* for such a component. On the other hand, they are not per se *security* experts. Their knowledge on specific attack routes and security vulnerabilities (e.g., SQL injection attacks) might be limited. Therefore, an internal expert's judgement of the vulnerability to *intentional attacks* might be less dependable.

B. Risk for whom?

In the case of the e-BioInfra, only the experts have been interviewed, which could indicate a bias towards internal information assets of the system.

The emphasis given to group members focuses the risk assessment on the internal stakeholders. MISRAM mainly measures the risk as perceived by the group and the group members. It does not take into account the risk for external parties such as users and, for example, patients. However, these kinds of risk are not at odds with each other. A low-risk science gateway is very likely to pose low risk for external parties as well.

C. Simplifications

MISRAM leaves out the percentage-based mapping of information assets used in the original DCRA. An information asset is linked to an IT component, or it is not. MISRAM also leaves out the propagation likelihood of worms and other attack vectors between nodes. Both simplifications come at a cost: our model is less precise than DCRA. The advantages are those of simple models: MISRAM is easy to work with and the reasoning and results are easy to communicate.

Note that MISRAM can easily be extended to include the full DCRA model, which could be benefitial to assess more complex systems or environments.

D. Connectivity

A component that is linked to many information assets will cause much risk because it is a hub; all information assets are in danger when such a hub is compromised. So, even a well administered server might cause quite some risk because so much depends on it. This could be called the "hub effect". This begs for the question of whether MISRAM measures risk or just measures the connectivity of a component. Is it reasonable to assign much risk to a hub, even if such a hub is well secured and shows little vulnerabilities?

One way to reduce the influence of this hub effect is to compensate a component for having many relations (connections) with information assets or with IT components. As an extreme example, one could divide the total risk caused by a component by the number of information assets to which the component is connected, resulting in the average risk per information asset caused by the component. A high average generated risk indicates that the component itself is vulnerable, regardless of the number of information assets to which it is connected.

Still we argue that IT components that are connected to many information assets are nearly always for a good reason on the priority list to harden against security problems. Such connected components are critical to the organisation. If the average risk per information asset caused by a component is low, but the total risk caused by the same component is high, then it means that the component has a very central function in the infrastructure. In such cases one should question whether the functions of the component could be split up among multiple components in order to isolate components and information flows.

E. CIA vs. business value and confidentiality

We have used the business value and confidentiality dimensions of the UK National Archives (UKNA) [7]. The UKNA model is easy to explain to non-experts and is also intuitive – factors that make it also easy to do a stakeholder survey on the information assets and to communicate the outcomes.

Another, very often used, way to describe security dimensions of an information asset is the "CIA" method, which stands for "confidentiality, integrity, and availability". The CIA approach is often used to describe the three main characteristics of information that can have value; they are also called the three main security categories [21, p. 5].

A mapping of the CIA to the UKNA model is possible. The 'C' from CIA is directly mapped to UKNA's confidentiality, while both 'I' and 'A' from CIA are mapped to UKNA's business value. Information with much business value should normally have good availability and integrity.

We chose to focus on integrity. The e-BioInfra is used to carry out scientific research, so the integrity of the data

processed is of paramount importance. High availability is currently not a major concern for us[5]. In other words, the correctness of the data processing is more important than the timeliness. For our approach, availability was out of scope, although it could easily be incorporated in MISRAM if needed.

F. Future work

We consider it questionable to compare different MISRAM risk assessments in different organisations because so many variables differ (such as the surveyed people and other social factors, external influences such as news and technical developments, and so on). Whether different organisations are comparable using MISRAM or an extension to our method could be determined by a follow-up project.

The right balance between stressing the importance of well-connected hubs and compensating for the hub effect is also a worthy subject of future research.

V. Acknowledgement

We thank the members of the e-Science group for their time and valuable suggestions during the development and application of this RA method. This work is partially supported by the COMMIT project "e-Biobanking with imaging for healthcare" funded by the Nederlandse Organisatie voor Wetenschappelijk Onderzoek (Netherlands Organisation for Scientific Research, NWO).

References

[1] S. Shahand, M. Santcroos, A. H. C. van Kampen, and S. D. Olabarriaga, "A grid-enabled gateway for biomedical data analysis," pp. 725–742.

[2] E. Deelman, D. Gannon, M. Shields, and I. Taylor, "Workflows and e-science: An overview of workflow system features and capabilities," *Future Generation Computer Systems*, vol. 25, no. 5, pp. 528–540, 2009.

[3] E. Mouw, G. van 't Noordende, A. H. van Kampen, B. Louter, M. Santcroos, and S. D. Olabarriaga, "Legal constraints on genetic data processing in European grids," in *HealthGrid Applications and Technologies Meet Science Gateways for Life Sciences*, 2012, pp. 49–58. [Online]. Available: http://www.booksonline.iospress.nl/Content/View.aspx?piid=30469

[4] ISACA, *CISA Review Manual 2006 (CISA - Certified Information Systems Auditor)*. Information Systems Audit and Control Association, 2006.

[5] E. Zambon, S. Etalle, R. J. Wieringa, and P. H. Hartel, "Model-based qualitative risk assessment for availability of IT infrastructures," *Software and Systems Modeling*, vol. 10, no. 4, pp. 553–580, 2011.

[6] I. Postelnik, "Chess: A valuable teaching tool for risk managers?" pp. 40–42, 2008.

[7] T. N. A. J. Riley), "What is an information asset?" 2011, factsheet. [Online]. Available: http://www.nationalarchives.gov.uk/

[8] G. Stoneburner, "NIST special publication 800-33 – underlying technical models for information technology security," National Institute of Standards and Technology, Computer Security Division, Tech. Rep., 2001. [Online]. Available: http://csrc.nist.gov/publications/nistpubs/800-33/sp800-33.pdf

[9] Cisco, "Data sheet technology application support – internal security posture assessment," 2004. [Online]. Available: http://www.cisco.com/application/pdf/en/us/guest/products/ps5619/c1262/cdccont_0900aecd800ce53a.pdf

[10] A. Morali, E. Zambon, S. Etalle, and P. L. Overbeek, "IT confidentiality risk assessment for an architecture-based approach," in *3rd IEEE/IFIP International Workshop on Business-driven IT Management*, 2008, pp. 31–40. [Online]. Available: https://ieeexplore.ieee.org/xpl/freeabs_all.jsp?arnumber=4540072

[11] J. yvind Aagedal, F. den Braber, T. Dimitrakos, B. A. Gran, D. Raptis, and K. Stlen, "Model-based risk assessment to improve enterprise security," in *EDOC '02. Proceedings. Sixth International Enterprise Distributed Object Computing Conference*, 2002, pp. 51–62.

[12] S. Strecker, D. Heise, and U. Frank, "RiskM: A multi-perspective modeling method for IT risk assessment," *Information Systems Frontiers*, vol. 13, no. 4, 2011. [Online]. Available: http://www.springerlink.com/content/j52k6071g4164q82/

[13] A. Morali, "IT architecture-based confidentiality risk assessment in networks of organizations," Enschede, 2011, IPA dissertation series; 2011-06.

[14] E. Mouw, "Data protection and privacy in escience," 2012. [Online]. Available: http://www.scriptiesonline.uba.uva.nl/425083

[15] M. Lankhorst and the ArchiMate team, "ArchiMate language primer," 2004. [Online]. Available: https://doc.novay.nl/dsweb/Get/Document-43839/

[16] U. Frank, "Multi-perspective enterprise modeling (MEMO) conceptual framework and modeling languages," in *HICSS. Proceedings of the 35th Annual Hawaii International Conference on System Sciences*, 2002, pp. 1258–1267.

[17] R. Likert, "A technique for the measurement of attitudes," *Archives of Psychology*, vol. 22, no. 140, pp. 1–55, 1932.

[18] H. L. Dreyfus and S. E. Dreyfus, *Mind over machine: The power of human intuition and expertise in the era of the computer*. Free Press, 1986.

[19] P. G. Schempp, "The stages of expertise," 2011. [Online]. Available: http://www.performancemattersinc.com/posts/stages-of-expertise/

[20] E. Mouw, "Internal report: Information security risk assessment of the eBioScience infrastructure at the AMC," eBioScience group, biolab, KEBB, Acedemic Medical Centre, University of Amsterdam, Tech. Rep., 2012, available on request. Contact the corresponding author.

[21] C. S. Division, "NIST special publication 800-53 rev. 3 – recommended security controls for federal information systems and organizations," National Institute for Standards and Technology, Tech. Rep., 2010. [Online]. Available: http://csrc.nist.gov/publications/nistpubs/800-53-Rev3/sp800-53-rev3-final_updated-errata_05-01-2010.pdf

[5] It is not a major problem if the e-BioInfra is offline for one or two days. The same seems to be the case of many peer research groups we know.