DRAFT

Full paper published in:
Metrology and Measurement Systems
Vol. 19, No. 2, 2012, pp. 269-282

Validation of Services Supporting Healthcare Standards Conformance

Janusz Górski[1,2], Aleksander Jarzębowicz[1,2], Jakub Miler[1,2]

1) Department of Software Engineering, Faculty of Electronics, Telecommunications and Informatics, Gdańsk University of Technology, G. Narutowicza 11/12, 80-233 Gdańsk, Poland (✉jango@eti.pg.gda.pl; olek@eti.pg.gda.pl; jakubm@eti.pg.gda.pl)
2) NOR-STA Project, Gdańsk University of Technology, G. Narutowicza 11/12, 80-233 Gdańsk, Poland, www.nor-sta.eu

Abstract

The paper presents the results of experimental validation of a set of innovative software services supporting processes of achieving, assessing and maintaining conformance with standards and regulations. The study involved several hospitals implementing the Accreditation Standard promoted by Polish Ministry of Health. First we introduce NOR-STA services that implement TRUST-IT methodology of argument management. Then we describe and justify a set of metrics aiming at assessment of the effectiveness and efficiency of the services. Next we present values of the metrics that were built from the data collected. The paper concludes with giving the interpretation and discussing the results of the measurements with respect to the objectives of the validation experiment.

Keywords: standards conformance, hospital quality management and monitoring, experimental validation, process metrics and measurements, NOR-STA services, TRUST-IT methodology.

1. Introduction

Achieving and demonstrating standards conformance is among the enablers of modern economy. Expected benefits from being conformant include improvements of quality, better position in the market, reduction of operating costs and others. An example of standards application are health services where standardisation helps in balancing the effectiveness of medical procedures with patient safety and privacy within the constraints of public funding. The role of standards and formal conformance certification becomes even more important when the responsibility for the service delivery is being moved from the public to private sector.

In Poland, the health sector involves (according to the official sources [1]) about 750 hospitals where approximately 80% are public and 20% are non-public institutions and the present tendency is that non-public share is growing. Some hospitals are formally certified against various international/national quality standards. The standard which presently gains dynamically growing popularity is the Accreditation Standard [2] promoted by National Center For Quality Assessment In Healthcare (Centrum Monitorowania Jakości w Ochronie Zdrowia, CMJ/NCQA) on behalf of Polish Ministry of Health. CMJ/NCQA was established by Polish Ministry of Health in 1994 to inspire and support improvement of quality of health services. The present status of certification in Polish healthcare is presented in Table 1.

It is expected that accreditation will play important role in healthcare quality assurance and that it will be recognized as a meaningful asset while applying for public funding. Introducing standards in a broad scale calls for effective and efficient tools support providing for automation of manual work, integration with business processes of the hospitals and facilitating decision making and resources allocation. A number of solutions (methods, approaches, frameworks) exist, however in most cases the support from IT tools is very limited. This limited support for standard conformance processes hinders massive application of standards [3].

Table 1. Present status of certification in Polish hospitals.

Type of certificate	# of certificates
Accreditation Standard	96
ISO 9001	332
ISO 14001	47
PN 18001	30
ISO 22000	6
ISO 27001	2
Total	513

The results of a detailed examination of some 30 software tools supporting standards conformance (from security standards domain which was our primary focus at the moment) are presented in [3], examples include: Callio Secura [4], CRAMM [5], Proteus [6], OJA [7], Tenrox Project Management [8] and NND Integrum [9].

The main problems and limitations include:

- focus on just one standard or a group of related standards,
- problems with widening the scope (addressing another standard), which requires significant development effort and new release of the software,
- lack of identifying and presenting interdependencies between standard's requirements and required conformance arguments,
- limited assessment functionality, mostly binary (yes-no) assessment, without representing uncertainty scales,
- limited support for maintaining conformance (as opposed to one time audit).

In this paper we introduce an innovative methodology [10-13] and the related set of software services (called NOR-STA services) that support application of standards. Then we describe a case study which is carried on since 2010, in which the services are deployed and used to support application of the Accreditation Standard. So far some 12 hospitals are involved, using the services at different stages of implementation of the standard. Next, we explain how we used Goal-Question-Metric (GQM) methodology [14] to construct a system of metrics aiming at assessment of effectiveness and efficiency of support given by NOR-STA services. The results of the measurements are presented in the subsequent section. In conclusions we interpret the measurement results and summarise them with respect to the measurement objectives.

2. TRUST-IT methodology and NOR-STA services

TRUST-IT [10-13] is an innovative approach to promoting trust by presenting in the cyberspace ‘live’ arguments integrated with the supporting evidence and providing means for assessing and visualizing the compelling power of the arguments. *Argument* is understood as an act of communication being an attempt to persuade someone of something, by giving reasons and/or evidence for accepting a particular conclusion. *Evidence* is a document in any form: text, graphics, image, web page, video, audio etc. which is used to demonstrate the facts

referred to in the argument. Integrating an argument with supporting evidence helps to make it more convincing. TRUST-IT introduces a model of an argument, a graphical language for expressing arguments and a technique for integrating arguments with the evidence. It also offers a general purpose argument appraisal mechanism based on Dempster-Shafer theory of beliefs and the corresponding mechanism of visualisation of the argument compelling power [15]. TRUST-IT methodology introduces the concept of a *trust case* which extends the concept of so called *safety case* commonly used in the safety-critical domain to justify safety properties of for instance, avionic, automotive, medical and military systems [16-18]. Trust cases are evidence based arguments that can be used to justify any selected property of a chosen object, including safety, security, privacy, conformance to predefined requirements and others. The argument model of TRUST-IT and the related graphical language for representing arguments are supported by a set of advanced tools that are offered to the users as software services..

TRUST-IT arguments were already developed to analyze safety, privacy and security issues of personalized health and lifestyle oriented services [19], monitoring of environmental risks [20] and support of standards conformance [3]. TRUST-IT is offered to its users by means of software services, called NOR-STA services. The services are deployed in accordance with the SaaS (Software as a Service) cloud computing model. They provide for representing and maintaining arguments, integrating them with the supporting evidence and assessing the compelling power of the arguments. Quality of service, in particular in relation to security is guaranteed by declaring and implementing an adequate security policy.

TRUST-IT approach is generic and can be applied in any context where evidence based argumentation brings added value to decision making processes and disputes. One of such application areas is standards conformance where a standard's user is obliged to construct and present an argument demonstrating conformance. While applied to standards conformance, TRUST-IT introduces additional, more specific concepts [3]. *Conformance argument template* is an argumentation structure derived from a standard. This structure is common for all conformance arguments related to the standard. *Conformance argument* is an argument which is constructed from the conformance argument template by extending the template with the evidence and possibly with more specific argumentation. *Conformance assessment* is an act of assigning appraisals to the conformance argument and its components to assess their 'compelling power'.

Fig. 1 explains how NOR-STA services are used to support processes related to achieving and assessing standards conformance. It distinguishes three processes: conformance argument template (CAT) development, conformance argument preparation (CAP) and conformance argument assessment (CAA). The notation is adopted from Eriksson-Penker business process modelling patterns [21]. Processes are activities (with possible complex internal structure) producing a specified output from a given input. In our diagrams processes are visualised as arrow-like shapes, while input and output resources are depicted as rectangles. An output of a given process can serve as an input to another process. Additional resources can be used by processes e.g. people performing activities, auxiliary physical or informational assets, various tools (including software) supporting activities. In our diagrams (Fig. 1, 2 and 3) two kinds of additional resources are present: people - depicted as a simplified human symbol, and software - depicted as an underlined circle. The resources are linked to the processes they support by dashed arrows.

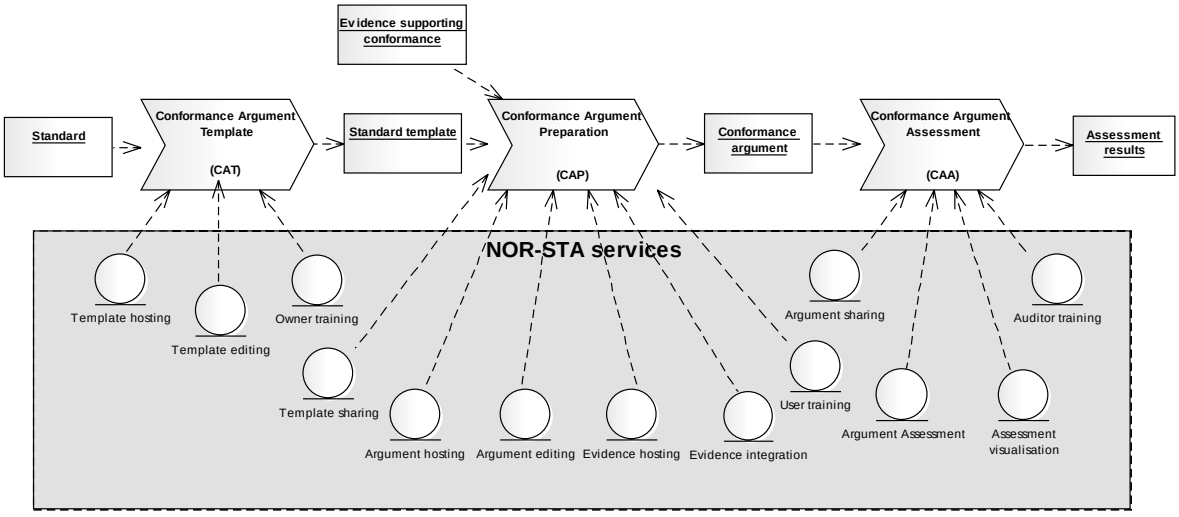

Fig. 1. Generic model of application of NOR-STA services.

3. Related work

Numerous solutions supporting application of standards are available in the market. They differ in their scope of applicability and in the extent of support provided. A thorough analysis which we have performed identified more than thirty approaches which could be compared to ours (the details are available in [3]). The criteria applied for comparisons included: universality (applicability to a wide range standards), scope of support, integration and visualization of the evidence, support for conformance evaluation and audit, on-line access in the Internet, support for conformance maintenance, and others (the details can be found in [3]). The main conclusion from the analysis is that presently no single generic solution exists capable of supporting application of a range of standards. Each of existing approaches targets at a small group of similar standards (mostly just one standard). For some standards, no dedicated solutions are available at all. With respect to this landscape, TRUST-IT and NOR-STA services offer a generic approach applicable to a broad range of standards and other normative documents, fully implemented in the cyberspace and supporting wide range of activities related to standards application, including achieving conformance, conformance assessment, and conformance maintenance.

4. Case Study: NOR-STA support for hospitals accreditation

Accreditation Standard [2] of Polish Ministry of Health (hereafter called A-standard) specifies quality requirements related to healthcare services and to well-being of a patient. The A-standard is based on internationally recognized best practices and recommendations issued by European Society for Quality in Healthcare [22].

The A-standard consists of 15 more specific standards (called *parts*) covering various aspects of processes taking place in a hospital: *Continuity of care, Patient's rights, Assessment of patient's state, Medical care, Infections' control, Treatments and anaesthetics, Pharmacotherapy, Laboratory, Image diagnostics, Nutrition, Quality improvement and patient's safety, General management, Human resource management, Information management and Environment management*.

An example requirement, borrowed from the *Continuity of care* part is as follows: *Hospital shall design and introduce patient admission procedures.* This requirement states that a hospital needs to design procedures describing both, planned and unplanned admissions

handling, with special regard to emergency situations. The requirement also provides more specific details like: the set of required examinations and documents, the scope of gathered personal identifiable information, the scope of medical history information gathered from the patient (in case of planned admission), aim of hospitalisation, the way of obtaining hospitalisation consent from the patient and others.

All 15 parts of A-standard include 221 requirements. Each requirement is associated with a list of tasks to be performed by an auditor while checking conformance and the related assessment scale (a distinction is being made between: *full conformance*, *partial conformance* and *lack of conformance*).

The process of using NOR-STA services to support A-standard conformance is a specialisation of the generic process presented in Fig.1 and is shown in Fig.2.

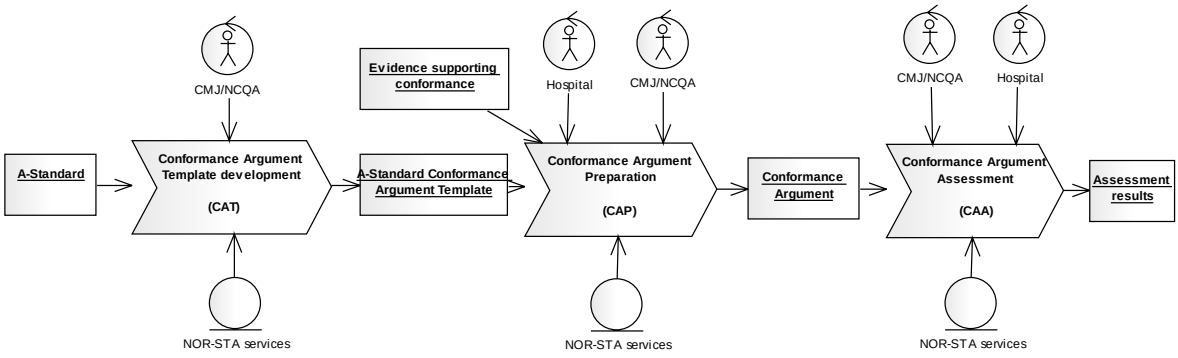

Fig. 2. The process of implementing NOR-STA services to support A-standard conformance

The details of CAT, CAP and CAA processes in healthcare are explained below.

4.1. CAT process

CAT resulted in templates of conformance arguments for all 15 parts of A-standard. The templates were built by CMJ/NCQA with NOR-STA support team providing consulting with respect to application of NOR-STA services and generic template structuring rules. The details of CAT process are presented in Fig.3.

CAT process consists of 4 sub-processes: *kick-off* (introducing the methodology and guidelines of conformance argument template construction), *design* (development of conformance argument templates for A-standard parts), *validation* (assessment of adequacy of the conformance argument templates) and *finalisation* (approval of the argument templates and their publication).

Design and validation sub-processes are repeated iteratively, where each iterated design step extends the conformance template by new requirements and possibly introduces corrections and improvements resulting from the preceding validation step. In our case study, three iterations were in effect, each iteration covering five additional parts of A-standard.

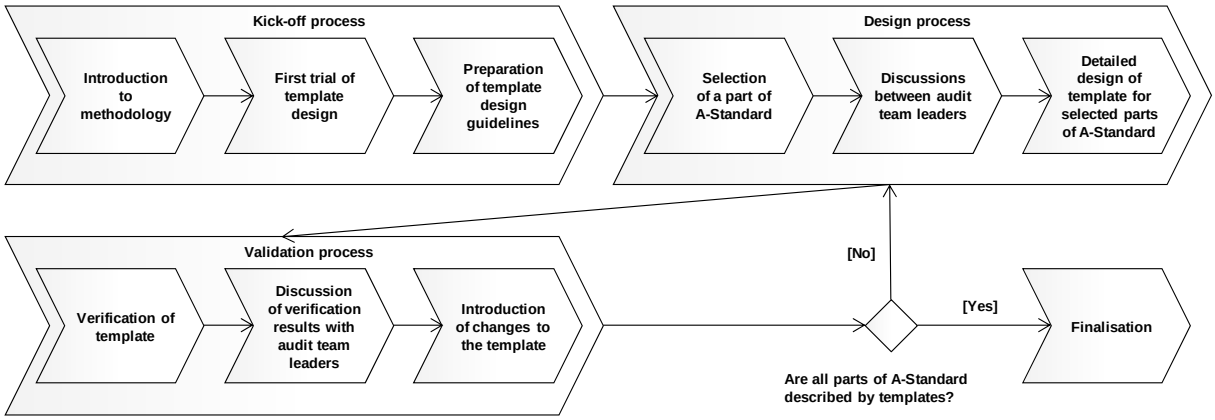

Fig. 3. CAT process structure.

4.2. CAP process

CAP process instances were performed by hospitals involved in the case study. Each hospital implementing A-standard was running a separate instance of CAP. In this paper we report the results obtained from implementing CAP in 6 different hospitals (denoted H1 to H6). Their characteristics are given in Table 2.

Table 2. Hospitals involved in the case study.

Hospital	Status	Size (# of wards)	Size (# of beds)
H1	public	15	387
H2	public	23	696
H3	public	16	465
H4	private	13	79
H5	public	17	405
H6	private	12	312

Each hospital was using a copy of the argument template of A-standard resulting from CAT process and converted it into a complete conformance argument. The structure of CAP process is shown in Fig.4.

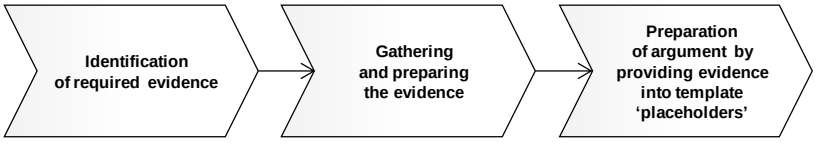

Fig. 4. CAP process structure.

Conformance arguments resulting from the instances of CAP processes were constructed by the involved hospitals with CMJ/NCQA providing consultancy concerning the template structure and its interpretation and the NOR-STA support team providing consultancy concerning application of NOR-STA services. The work of hospitals mainly involved producing and/or collecting pieces of the evidence and integrating them with the conformance argument template.

The implemented CAP processes differed depending on how advanced was the related hospital in its preparation to the accreditation. For 3 hospitals, NOR-STA services were used in retrospective mode: reconstructing the conformance argument after the accreditation was already granted. For 3 other hospitals, NOR-STA services were used in pro-active mode: supporting construction of the conformance argument while preparing for accreditation.

4.3. CAA process

CAA processes were performed taking as an input the conformance arguments resulting from the instances of CAP process. The CAA used the 3-element assessment scale defined by the A-standard. The details of CAA process are given in Fig. 5.

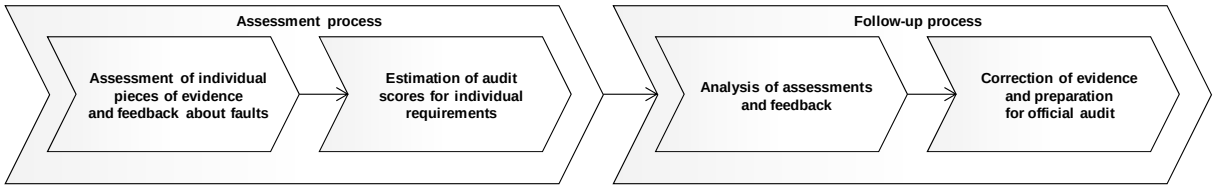

Fig. 5. CAA process structure

The CAA process distinguishes between the assessment and the follow-up phases where the latter involves implementation of improvements and corrections suggested by the former one. These preliminary assessments resulted in detection of several deficiencies in the presented evidence and the resulting improvements.

5. Establishing a metrics system

The case study described in the previous section was used to collect data which were then used to validate the NOR-STA services. The validation criteria and the related set of metrics were identified using a common Goal-Question-Metric (GQM) methodology [14]. GQM proposes systematic approach to derivation of metrics starting from an explicit statement of the measurement goal. The link between the goal and the metrics is established and controlled by the intermediate layer of questions, answering to which helps in deciding to which extent the selected metrics support the stated goal. Then, for each metric, the data collection mechanisms are defined and implemented.

Following the GQM method, the overall goal for our validation experiment was expressed as follows (this is an instantiation of the standard GQM goal definition template): *Analyze* ***NOR-STA services*** *for the purpose of* ***improvement*** *with respect to* ***effectiveness of achieving and assessing conformance*** *from the viewpoint of* ***standard's owner, user and auditor*** *in the context of* ***application of A-standard***.

At a lower decomposition level, we identified a set of questions reflecting the above goal.

Table 3. Questions related to the measurement goal

ID	Question text
Q1	What is the effectiveness of NOR-STA services in CAT process?
Q2	What is the effectiveness of NOR-STA services in CAP process?
Q3	What is the effectiveness of NOR-STA services in CAA process?
Q4	What are the benefits of NOR-STA services for an institution applying the standard?
Q5	What is the effectiveness of software tools implementing NOR-STA services?

The third level of GQM decomposition involves identification of metrics which are used to answer a particular question. The metrics are associated with corresponding Data Collection Mechanisms (DCM). The following DCMs were used:

- Questionnaires (QQ) – data collected by using questionnaires filled by the users;
- Automatic measurements (AM) – data collected automatically by corresponding software implemented mechanisms;

– Document reviews (DR) – data collected from the documents subjected to manual review;
– Experiment result (ER) – data collected by interpreting the results of experiments;
– Manual inspection (MI) – data collected by manual inspection of a corresponding object.

The metrics and the corresponding data collection mechanisms are given in Tables 4 to 8.

Table 4. Metrics associated with Q1.

ID	Metrics description	DCM
M1.1	Size of the standard	DR
M1.2	Number of requirements of the standard	DR
M1.3	Possibility of expressing the standard as a conformance template	ER
M1.4	Coverage of the standard by the template	DR
M1.5	Number of people involved in CAT process	QQ
M1.6	Effort on CAT process	QQ
M1.7	Effort per a requirement	AM
M1.8	Number of problems with application of NOR-STA services to CAT process	QQ
M1.9	Effort on learning to apply the NOR-STA services to CAT process	QQ

Metrics M1.1 and M1.2 were expressed as numeric values derived from the original text of the A-standard. M1.3 was measured in the binary scale [*yes, no*]. M1.4 was expressed as the percentage of requirements included in the template. Metrics M1.5, M1.6, M1.8 and M1.9 were derived from the questionnaire filled by CMJ/NCQA (the developer of the template). M1.7 was calculated from M1.6 and M1.2.

Table 5. Metrics associated with Q2.

ID	Metrics description	DCM
M2.1	Completeness of the references	QQ
M2.2	Accessibility to the referred evidence	QQ
M2.3	Separation of additional evidence from the template references	QQ
M2.4	Integrity of the template (number of departures from template structure)	QQ
M2.5	Appropriateness of the evidence	QQ
M2.6	Substitution of the evidence by a declaration of existence	QQ
M2.7	Total number of arguments developed	MI
M2.8	Total number of evidence files in the repositories	MI
M2.9	Average duration of a session with NOR-STA services	AM

Metrics M2.1 to M2.6 verify the structural correctness of the arguments. They were measured with help of a questionnaire filled by a CMJ/NCQA auditor, who answered each *yes/no* question and in case of a negative answer provided an explanation. Metrics M2.7 and M2.8 were measured by inspecting the NOR-STA services software. M2.9 was measured in an automatic way from the data contained in a log associated with NOR-STA software.

Table 6. Metrics associated with Q3.

ID	Metrics description	DCM
M3.1	Number of arguments assessed	MI
M3.2	Number of auditors involved	AM
M3.3	Number of assessments	AM
M3.4	Total duration of assessment sessions with NOR-STA services	AM
M3.5	Average duration of an assessment session with NOR-STA services	AM

Metric M3.1 was measured by direct examination of the conformance arguments in NOR STA services software. Metrics M3.2 to M3.5 were measured in an automatic way from the data contained in a log built into the NOR-STA services.

Table 7. Metrics associated with Q4.

ID	Metrics description	DCM
M4.1	Benefits to the standard's users	QQ
M4.2	Benefits to the auditors	QQ

Metrics M4.1 and M4.2 where further decomposed as shown in Table 7a and Table 7b.

Table 7a. Submetrics of M4.1.

ID	Metrics description	DCM
M4.1.1	Time saving – shorter process	QQ
M4.1.2	Effort saving	QQ
M4.1.3	Cost saving	QQ
M4.1.4	Arrangement of the evidence	QQ
M4.1.5	Security of the evidence	QQ
M4.1.6	Distribution and promotion of a standard within the user's organization	QQ
M4.1.7	Promotion of the conformance in the market	QQ
M4.1.8	Preparation to the audit	QQ
M4.1.9	Maintenance of the conformance	QQ
M4.1.10	Clear information on conformance to the management	QQ

Table 7b. Submetrics of M4.2.

ID	Metrics description	DCM
M4.2.1	Overview of the evidence prepared by an institution to audit	QQ
M4.2.2	Preparation to the official audit	QQ
M4.2.3	Overview of the evidence coverage	QQ
M4.2.4	Justification of the auditor's decisions	QQ
M4.2.5	Matching of the assessment criteria among the auditors	QQ
M4.2.6	Training of the auditors based on arguments and assessments	QQ
M4.2.7	Consulting the hospitals based on the assessments	QQ
M4.2.8	Using the evidence in the appeal case	QQ

For each benefit from M4.1 and M4.2 we measured:

a) business value for the standard's user (hospital) or the auditor,

b) support from NOR-STA services in achieving the business value.

Metrics M4.1 were measured based on the answers to a questionnaire for standard's users (hospitals) and the auditors. Metrics M4.2 were measured from the answers to a questionnaire for the auditors only. Business value used a linguistic scale [*minor, moderate, major*], while the support from NOR STA services was measured in a linguistic scale [*none, minor, major*].

Table 8. Metrics associated with Q5.

ID	Metrics description	DCM
M5.1	Effectiveness of the dedicated assessment scale	QQ
M5.2	Effectiveness of the user interface in Polish	QQ
M5.3	Effectiveness of the simplified evidence attachment	QQ
M5.4	Effectiveness of the comments to the assessments	QQ
M5.5	Effectiveness of copying the fields between nodes	QQ
M5.6	Usability of the NOR-STA services user interface	QQ
M5.7	Reliability of the NOR-STA services	QQ
M5.8	Performance of the NOR-STA services	QQ
M5.9	Availability of the NOR-STA services in the Internet	AM

For each metric from M5.1 to M5.5 we measured:

a) business value for the standard's user (hospital) or the auditor,

b) support from NOR-STA services in achieving the business value.

Metrics M5.1 to M5.8 were measured in the scale [*1, 2,...,5*] (1 - the worst, 5 - the best) based on the answers to a questionnaire distributed among the auditors and the quality managers of the hospitals. M5.9 was measured by an external monitoring service [23] as a percentage of time in which the NOR-STA services server was reachable from the Internet.

We derived the metrics systematically following the GQM methodology and starting from the measurement goal. The selection of metrics and the data collection mechanisms were consulted with the CMJ/NCQA representatives who were representing the auditor's and standard owner's viewpoint. Automatic measurements (AM), document reviews (DR), experiment results (ER) and manual inspections (MI) data collection mechanisms were executed without external help, whereas the questionnaires (QQ) were designed in collaboration with CMJ/NCQA. The questionnaires were then filled by the quality managers of the hospitals involved in the experiments. This took place during the NOR-STA Project Advisory Committee meeting in December 2011. Then, the data from the questionnaires were extracted, processed and the corresponding metrics were calculated.

6. Measurements results

The values of the metrics related to efficiency of argument template development, completeness and integrity of the resulting conformance argument templates, and efficiency of argument preparation and assessment are presented in Tables 10 to 13.

Table 10. Metrics related to efficiency of argument template development (CAT).

Metric	Value
M1.1	129 A4 pages
M1.2	221
M1.3	Yes
M1.4	100%
M1.5	4
M1.6	393 person-hours
M1.7	1.78 person-hours/requirement
M1.8	16 problems
M1.9	28 person-hours

Table 11. Metrics related to argument completeness and integrity (collected from hospitals H1 and H2).

Metric	H1	H2
M2.1	No, 15% missing	No, 28% missing
M2.2	No, 1 problem	No, 1 problem
M2.3	Yes	Yes
M2.4	Yes	Yes
M2.5	No, 4 problems	Yes
M2.6	Yes, 16% of evidence	Yes, 28% of evidence

Table 12. Metrics related to efficiency of argument preparation (CAP).

Metric	Value
M2.7	6
M2.8	1075
M2.9	28 min.
M5.9	99.6%

Table 13. Metrics related to efficiency of argument assessment (CAA).

Metric	Value
M3.1	4
M3.2	3
M3.3	905
M3.4	32.75 hours
M3.5	1.25 hours

Effectiveness of the support delivered by NOR-STA services to the institutions applying standards was measured by the metrics related to questions Q4 and Q5.

Fig. 6 presents the values of metrics for Q4 in group M4.1 ordered by descending business value and then (for the metrics of equal business value) by descending NOR-STA support. The scales [*minor, moderate, major*] and [*none, minor, major*] were converted into the scale [*1, 3, 5*] and the average values were calculated.

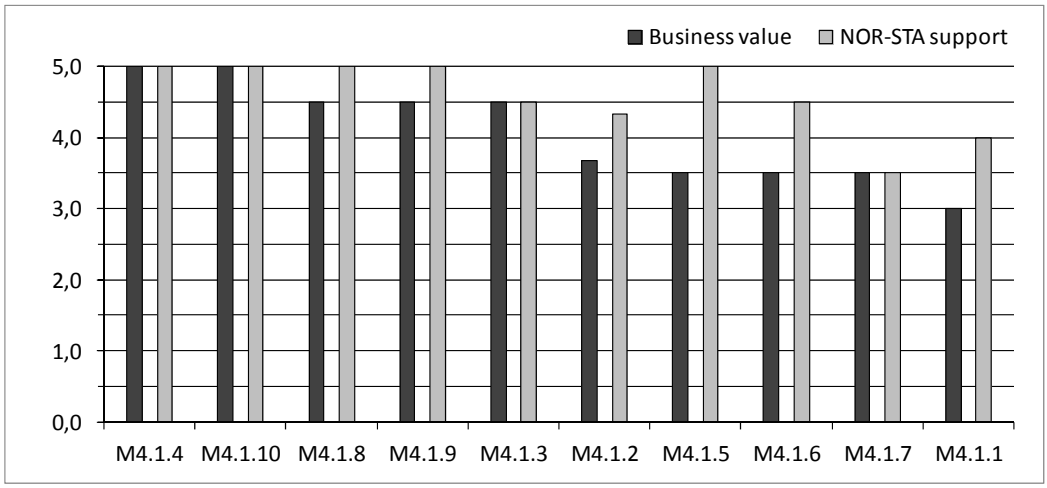

Fig. 6. Effectiveness of support for achieving conformance (metrics M4.1.1 to M4.1.10)

Fig. 7 presents the values of metrics for Q4 in group M4.2 ordered by descending business value and then (for the metrics of equal business value) by descending NOR-STA support. The scales [*minor, moderate, major*] and [*none, minor, major*] were converted into the scale [*1, 3, 5*].

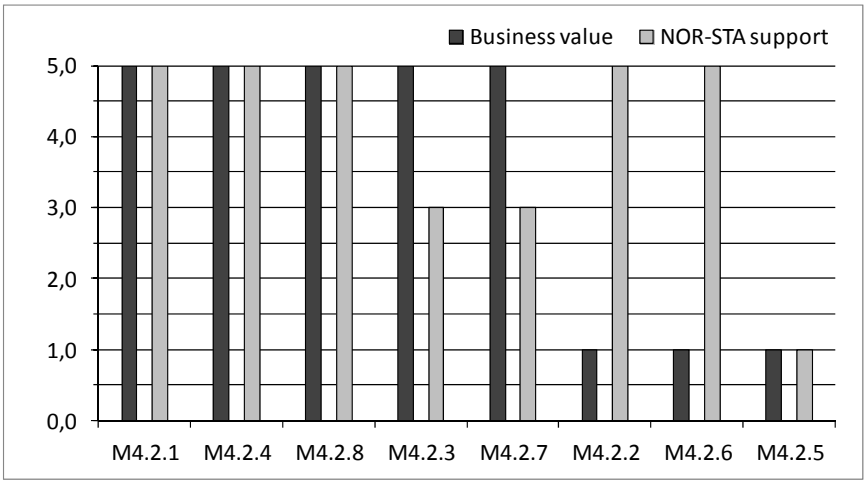

Fig. 7. Effectiveness of support for assessing conformance (metrics M4.2.1 to M4.2.8)

Fig. 8 presents the values of metrics M5.1 to M5.5 ordered by descending business value and then by descending NOR-STA support. In total, 3 questionnaires were collected and the average of the answers was calculated as the values of the metrics.

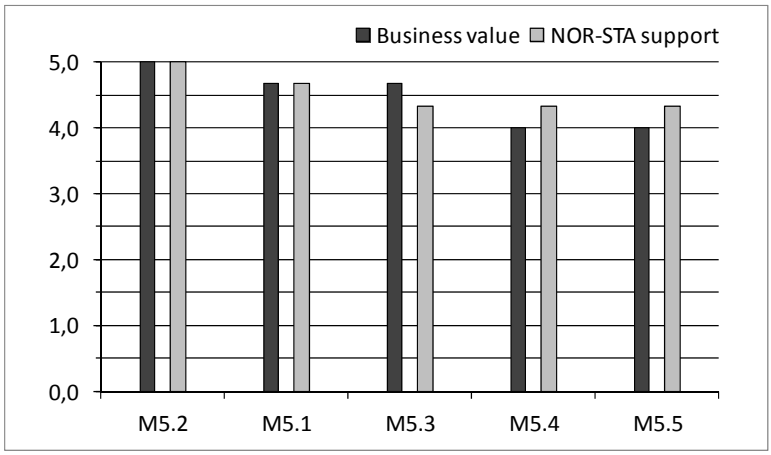

Fig. 8. Effectiveness of the software implementing NOR-STA services (metrics M5.1 to M5.5)

Concerning quality of the NOR-STA services, the values of the corresponding metrics were as follows: M5.6=4.0, M5.7=4.35, M5.8=4.65 in the [*1,2,...,5*] scale.

6. Conclusions

TRUST-IT is a generic methodology aiming at development, maintenance, appraisal and communication of arguments. One of important application domains for TRUST-IT is achieving and assessing standards conformance. For this purpose, TRUST-IT is facilitated by NOR-STA services available in the Internet. NOR-STA services are generic and can be applied to any standard. The concept of conformance argument template provides for taking into account the particularities of a given standard and at the same time introduces a common structure of conformance arguments for that standard.

The paper presented the results of experimental validation of NOR-STA services aiming at assessing their effectiveness and efficiency. The subject of research was the Accreditation Standard being presently proposed for Polish hospitals. The paper described the scope of the case study and then presented how the set of adequate metrics has been selected in a systematic way following the GQM approach. The metrics were associated with corresponding data collection mechanisms and the data were collected during the study that spanned over some 18 months. The results of the study can be summarized as follows.

- It was feasible to develop a conformance template for a not-trivial standard and cover all its requirements (M1.1, M1.2, M1.3, M1.4).
- Conformance template development requires considerable effort (but it is done only once for a given standard) (M1.5, M1.6, M1.7).
- NOR-STA services are easily adopted by template developers – the learning effort was small comparing to the total effort of the template development (M1.6, M1.9).
- Different hospitals were able to build their conformance arguments of non-trivial size, although with some minor problems (M2.1 to M2.6); Six independent conformance arguments have been developed involving more than 1000 evidence files (M2.7, M2.8).
- Availability of NOR-STA services was sufficient (M2.9, M5.9).
- It was feasible to assess several non-trivial conformance arguments (M3.2, M3.3); the argument appraisal services were used by several auditors (M3.1, M3.3, M3.4).

- The assessment was reasonably fast (around 2 minutes per requirement); however, a break was needed after each 1.25 hours long assessment session (M3.3, M3.4, M3.5).
- NOR-STA services support best the most beneficial aspects of the conformance achieving process: evidence management and decision support (Fig. 6: M4.1.4, M4.1.10). In general, the support is adequate to the business value declared by hospitals (for each business goal, the support is not lower than the declared business value) (Fig. 6).
- From the audit viewpoint, the services need improvement (2 out of 5 goals have inadequate support with respect to the declared business value) (Fig. 7: M4.2.3, M4.2.7).
- The software implementation is well adjusted to users' expectations (Fig.8); The translation of the user interface to Polish was greatly appreciated by the users (M5.2).
- Efficiency, reliability and availability of the services were acceptable to the users (M5.7, M5.8, M5.9).

The raw data collection mechanisms and the related metrics are embedded in the process of incremental deployment of NOR-STA services for different standards used in different application domains. The metrics provide feedback which is actively used to improve the services and their usage procedures in subsequent increments. The scope of the metrics covers different aspects, including the business value for the end users – the factor of critical importance for determining an adequate model of introducing the services to the market.

GQM methodology appeared to be very useful in defining the scope of metrics and the scope of collected data in a systematic way, starting from the explicitly stated objective. This provides for ending with a set of metrics which is minimal in a sense that it is sufficient for meeting the objective and does not include metrics which are redundant. This limits the effort needed for sometimes very costly process of data collection. Nevertheless, the relationship between the layers of GQM (the goal, the questions, the metrics, the data collection mechanisms) still heavily depends on the expert judgement and often needs trade-offs between what one would like to measure and what is possible to be measured within the given constraints (time, effort, availability of raw data and so on).

From our experience of working with actual standards we conclude that NOR-STA services could also be used by Standard Bodies during the design of a new standard. In such case, representing the standard as a conformance template would help in detecting possible deficiencies of the standard's structure and content, such as ambiguities and inconsistencies of the requirements, as well as would help in presenting the standard in a better structured form, facilitating direct application of the standard.

Acknowledgements

This work was partially supported by the NOR-STA project co-financed by the European Union under the European Regional Development Fund within the Operational Programme *Innovative Economy* (grant # UDA-POIG.01.03.01-22-142/09-03). Contribution to technical development of NOR-STA services by Michał Witkowicz, Jakub Czyżnikiewicz and Patryk Jar and cooperation of Michał Bedlicki from CMJ/NCQA in Kraków are also acknowledged.

References

[1] Central Statistical Office, Concise Statistical Yearbook of Poland 2011, Statistical Publishing Establishment, ISSN 1640-3630, 259.

[2] Accreditation Standards, National Centre for Monitoring Quality (CMJ/NCQA), Krakow 2009.

[3] Cyra, L., Górski, J. (2011). SCF - a Framework Supporting Achieving and Assessing Conformity with Standards, Computer Standards & Interfaces, Elsevier, 33, 80-95.

[4] Callio Technologies home page, http://www.callio.com, (visited on 4 April 2012).

[5] CRAMM home page, http://www.cramm.com, (visited on 4 April 2012).

[6] Information Governance Limited home page, http://www.infogov.co.uk, (visited on 4 April 2012).

[7] Prim Information Technology home page, http:// www.prim-it.pl, (visited on 4 April 2012).

[8] Tenrox Project Management Software home page, http:// www.tenrox.com, (visited on 4 April 2012).

[9] Tkomp home page, http://www.tkomp.pl/, (visited on 4 April 2012).

[10] Górski, J. (2005). Trust Case – a case for trustworthiness of IT infrastructures. In *Cyberspace Security and Defense: Research Issues, NATO Science Series II: Mathematics, Physics and Chemistry*, 196, Springer-Verlag, 125-142.

[11] Górski, J., Jarzębowicz, A., Leszczyna, R., Miler, J., Olszewski, M. (2005). Trust Case: Justifying Trust in IT Solution. *Elsevier, Reliability Engineering and System Safety,* 89, 33-47.

[12] Górski, J. (2007). Trust-IT – a framework for trust cases , Workshop on Assurance Cases for Security - The Metrics Challenge. *Proc. of DSN 2007*, Edinburgh, UK, 204-209.

[13] NOR-STA project Portal, *www.nor-sta.eu* (2012).

[14] van Solingen R., Berghout E., (1999) The Goal/Question/Metric method: A practical guide for quality improvement of software development, *McGraw-Hill Publishing Company*, England.

[15] Cyra, L., Górski, J. (2011). Support for Argument Structures Review and Assessment. *Reliability Engineering and System Safety, Elsevier*, 96, 26-37.

[16] Emmet L., Guerra S., (2005) Application of a Commercial Assurance Case Tool to Support Software Certification Services, Proceedings of the 2005 Automated Software Engineering Workshop on Software Certificate Management (SoftCeMent'05), Association for Computing Machinery, New York, pp. 51-55.

[17] Kelly T., Weaver R., (2004) The Goal Structuring Notation - A Safety Argument Notation, Proceedings of the Dependable Systems and Networks Workshop on Assurance Cases.

[18] Kelly T., (2003) Managing Complex Safety Cases, Proceedings of 11th Safety Critical Systems Symposium, Springer-Verlag.

[19] Górski, J., Jarzębowicz, A., Miler, J. (2008). Arguing trustworthiness of e-health services with the Trust-IT framework. *25th Anniversary Healthcare Computing: Invitation to the Future: Conference & Exhibition (HC 2008)*, Harrogate.

[20] Proc. of the Workshop on Selected Problems in Environmental Risk Management and Emerging Threats, June 2009, Gdansk, Poland (http://kio.pg.gda.pl/ERM2009/)

[21] Eriksson, H.-E., Penker, M. (2000). Business Modeling with UML. Business Patterns at Work, J Wiley.

[22] European Society for Quality in Healthcare, *www.esqh.net* (2012).

[23] Monitoring of e-services and websites, monit24.pl (2012).