HAL
open science

Limites de l'évaluation d'un système d'information : une analyse fondée sur l'expérience pratique

Káthia Marçal de Oliveira, Virginie Thion, Sophie Dupuy-Chessa, Marie-Pierre Gervais, Samira Si-Said Cherfi, Christophe Kolski

To cite this version:

Káthia Marçal de Oliveira, Virginie Thion, Sophie Dupuy-Chessa, Marie-Pierre Gervais, Samira Si-Said Cherfi, et al.. Limites de l'évaluation d'un système d'information : une analyse fondée sur l'expérience pratique. 30ème Conférence INFormatique des ORganisations et Systèmes d'Information et de Décision (INFORSID), May 2012, Montpellier, France. pp.395-410. hal-00953467v2

HAL Id: hal-00953467

https://hal.science/hal-00953467v2

Submitted on 14 Feb 2024

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

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

Limites de l'évaluation d'un Système d'Information : une analyse fondée sur l'expérience pratique

Káthia Marçal de Oliveira*, Virginie Thion, Sophie Dupuy-Chessa***, Marie-Pierre Gervais****, Samira Si-Said Cherfi**, Christophe Kolski***

** Laboratoire LAMIH – UMR CNRS 8201*
Université de Valenciennes - F-59313 Valenciennes cedex 9
Kathia.oliveira@univ-valenciennes.fr, Christophe.Kolski@univ-valenciennes.fr

*** Laboratoire CEDRIC*
Conservatoire National des Arts et Métiers- F-75141 PARIS cedex 03
Virginie.Thion@cnam.fr, samira.cherfi@cnam.fr

**** Laboratoire d'Informatique de Grenoble*
Université de Grenoble - 385, av. de la Bibliothèque, 38400 Saint-Martin d'Hères
Sophie.Dupuy@imag.fr

***** Université Paris Ouest Nanterre et Laboratoire d'Informatique de Paris 6*
Boîte 169 - 4 Place Jussieu, 75252 Paris Cedex 05
Marie-Pierre.Gervais@lip6.fr

RÉSUMÉ. *Un grand nombre de définitions du système d'information (SI) ont été proposées dans la littérature. Il est maintenant bien accepté par les communautés manipulant ce concept que le système d'information combine un système d'information numérique et des activités humaines afin de servir de support aux branches opérationnelles, à la gestion et à la prise de décision dans l'entreprise. Assurer la qualité du SI est donc crucial. Cependant, évaluer cette qualité est une tâche multidimensionnelle complexe, concernant différentes ressources (processus, données, artéfacts, interfaces, …), considérant différentes caractéristiques (utilisabilité, fonctionnalités, …) et mettant en œuvre différentes méthodes pour effectuer l'évaluation. Après plusieurs années d'expérience dans l'évaluation des SI, nous résumons dans ce papier les leçons que nous avons tirées des évaluations que nous avons menées, et exhibons les limites de l'évaluation d'un SI. Nous finissons le papier par une réflexion sur l'avenir de l'évaluation d'un SI.*

ABSTRACT. *Despite the several different definitions about information system (IS), it is commonly accepted that it combines information technology and people's activities to support operations, management, and decision making. To assure its quality is, therefore, crucial for the success of its real adoption. However, evaluating an IS is a multi-dimensionnal and not trivial task, involving different entities (process, data, artifacts, interface, …), considering different characteristics (functionality, usability, ..) and different*

methods to support the evaluations. After years of experience in evaluation of IS, we synthetize on this paper our lessons learned, what we can do and what are the limits of an evaluation. We finish the paper, with a look forward about the future on evaluation of IS.

MOTS-CLÉS : *qualité, système d'information, évaluation*

KEYWORDS: *quality, information system evaluation*

1. Introduction

Développer un système d'information (SI) robuste et réactif pour appuyer la planification formelle et le contrôle des processus organisationnels est de plus en plus important (Irani, 2002). Défini comme étant « un ensemble organisé de ressources (matériel, logiciel, personnel, données, procédures) permettant d'acquérir, de traiter, de stocker, de communiquer des informations (sous forme de données, textes, images, sons, etc.) dans les organisations » (Reix, 2002), un SI ne peut être utile que s'il est capable de fournir l'ensemble des informations et des services pour lesquels il a été conçu. Évaluer la qualité des SI devient dans ce contexte un enjeu stratégique, qui a de plus un impact direct sur son utilisation et son adoption. Cependant, l'évaluation d'un SI n'est jamais une tâche facile.

De nombreux cadres pour l'évaluation de la qualité d'un SI, visant à intégrer différents points de vue de celui-ci, ont été proposés (Comyn-Wattiau *et al.*, 2010). Par exemple (Stylianou *et al.*, 2000) propose des dimensions suivantes : la qualité de l'infrastructure, la qualité logicielle, la qualité des données, la qualité de l'information, la qualité administrative et la qualité des services fournis par le système au client. Chrissis *et al. (*2003) considèrent trois dimensions (les personnes, les procédures utilisés et les outils et hardware) qui pour travailler ensemble ont besoin d'un processus logiciel, décrivant les activités et artefacts générés. En considérant ces travaux et notre propre expérience nous allons considérer dans cet article les dimensions suivantes : la qualité des processus utilisés dans une organisation, la qualité des artefacts générés pendant le développement, la qualité des données et, finalement, la qualité perçue par l'utilisateur en utilisant le SI, c'est-à-dire, la qualité des interfaces finales du SI.

Après plusieurs années d'expérience dans l'évaluation des SI, nous proposons de résumer dans ce papier les leçons tirées des évaluations que nous avons menées sur des processus, des artefacts, des interfaces et des données. Nous organisons ces leçons en exposant, à partir d'une brève présentation des travaux existants, les difficultés et limites des évaluations que nous avons pu constater en pratique.

L'article est organisé de la façon suivante. Les sections 2, 3, 4 et 5 présentent une analyse de l'évaluation des processus, artefacts, interfaces et données respectivement. La section 6 présente une synthèse de ces analyses, avant de dresser nos conclusions et quelques perspectives du domaine dans la section 7.

2. L'évaluation des processus

2.1. *Travaux existants*

Un « Processus » est défini comme « un ensemble d'activités corrélées ou interactives qui transforme les entrées en sorties » (ISO 9000, 2000). Afin d'améliorer la performance des projets de développement des SI, les organisations cherchent à investir dans la définition et l'amélioration de leurs processus (Raffo, 2005). Dans cet article, nous considérons de manière indistincte tous les types de processus (métier, développement etc) Toutefois, les organisations éprouvent des difficultés à gérer leurs processus. Comment définir au mieux les processus logiciels ? Y-a-t-il de vrais bénéfices dans l'utilisation d'un processus pour le développement d'un SI ? Les produits sont-ils de meilleure qualité ? Comment améliorer le processus défini ? Les recherches existantes sur l'évaluation des processus visent à répondre à ces questions et d'autres dérivées.

Pour appuyer la définition des processus logiciels, plusieurs normes (e.g., ISO/IEC 12207 (ISO/IEC, 2008), ISO/IEC 15504(ISO/IEC, 2004)) et modèles de références de processus (e.g. CMMI-*Capability Maturity Model Integration* (SEI, 2011)) ont été proposés et sont utilisés dans le monde académique et l'industrie. Ils décrivent des procédures génériques qui doivent être intégrées dans la définition des différents processus d'une organisation, mais ne considèrent pas la culture spécifique de chaque entreprise. De nombreux travaux sont proposés dans ce domaine (e.g. Staples et Niazi, 2010 ; Hwang, 2009 ; Horie *et al.*, 2009). Nous-mêmes avons utilisé ces standards pour la définition des processus (Machado *et al.*, 2000) et nous avons réalisé une étude sur les difficultés et les facteurs de succès avec l'utilisation de processus basée sur CMMI (Rocha *et al.*, 2006). Pour la définition des processus, un modèle de processus (p.e., modèle de processus métier) peut être élaboré et sa qualité doit aussi être assurée comme discuté dans la session suivante.

Pour assurer la qualité des produits (artefacts) générés avec les processus, des processus spécifiques d'évaluation sont établis et des mesures de qualité sont utilisées. Les travaux existants montrent comment définir, collecter et interpréter ces mesures. Dans ce contexte, nous avons exploré différentes pistes. Nous avons réalisé une large étude de la littérature pour définir un ensemble de mesures, procédures de collecte et d'analyse pour l'évaluation de la performance de processus logiciel (Monteiro et Oliveira, 2011). Nous avons également mené des recherches empiriques pour évaluer si l'utilisation de guides (modèles) de processus a des effets sur la qualité des produits finaux. L'expérience a permis de montrer que les résultats obtenus sont directement liés au cas d'étude considéré (Almeida *et al.*, 2010). Une réponse générale s'avère donc impossible. Une autre étude empirique nous a révélé qu'il existe une grande différence entre les attentes des clients des SSII (qui utilisent des processus basés sur CMMI) en termes de qualité et la perception de la qualité du service fourni (Santos et Oliveira, 2009). Enfin, nous avons étudié l'évaluation de l'impact des déviations générées par les agents exécutant le processus *a priori* conformément à son modèle. Or il est reconnu que les agents ne respectent pas le

modèle à 100% (Lanubile *et al.*, 2000). Il est alors nécessaire de détecter les déviations, et de les traiter c'est-à-dire analyser le risque qu'elles compromettent le bon déroulement de la suite du projet (Almeida *et al.*, 2011).

Finalement, pour améliorer continuellement la qualité de processus, des travaux promeuvent l'utilisation de modèle d'amélioration comme le paradigme QIP (Quality Improvement Paradigm) qui utilise Goal-Question-Metric (Basili *et al.*, 1994) et les méthodes DMAIC (Define-Measure-Analyse-Improve-Control) (De Feo and Barnard, 2005). Considérant qu'une amélioration implique de modifier le processus, nous sommes en train d'étudier comment évaluer *a priori* l'impact d'éventuelles modifications que nous envisageons sur un processus. Ce travail est mené en lien avec le traitement des déviations. Il s'agit d'identifier si une déviation peut être intégrée dans le processus, ce qui implique sa modification. Ceci n'est acceptable que si la modification ne compromet pas la suite du processus (par exemple, ne fait pas apparaître un risque d'interblocage entre activités). Dans d'autres travaux, nous nous sommes intéressés à l'évaluation (sur un cycle DMAIC complet) de la qualité d'un processus métier de pilotage dont tous les acteurs étaient des acteurs humains (Grim-Yefsah *et al.*, 2010).

2.2. *Limites et difficultés de l'évaluation de processus*

À partir de nos connaissances de la littérature et de nos expériences de recherches présentées dans la section précédente, nous identifions les difficultés et limites pour l'évaluation de processus, citées ci-dessous :

- **grand nombre de mesures de processus** – les travaux précités mentionnent plus de 500 mesures de processus logiciels différentes (Monteiro et Oliveira, 2011). Choisir l'ensemble de mesures à évaluer est complexe ;
- **difficulté de collecter et interpréter les mesures** – les procédures de collecte doivent être intégrées dans les processus logiciels. Si ces mesures sont automatisables ou si l'entreprise dispose d'un logiciel pour la gestion de processus, cette intégration est plus facile et fiable que lorsque les mesures dépendent d'activités manuelles (comme la collecte lors de réunions d'inspection). De plus, afin que les mesures soient valides et utiles, un historique des données est nécessaire. Enfin, une fois celles-ci collectées, l'interprétation dépend souvent d'une analyse d'experts. Notre expérience montre que les résultats sont souvent approximatifs et que les seuils ne suffisent pas, car le contexte du projet (taille, équipe de développement, culture, pression des agendas, etc.) influence l'interprétation finale ;
- **difficulté d'évaluer à l'avance si la modification doit être faite** – l'impact d'une déviation peut être sans risque pour la suite du processus, et dans ce cas, le processus peut être modifié. Mais si la déviation apporte un risque d'incohérence, il faut alors préserver le processus en interdisant la déviation.

- **choix de modélisation et choix du formalisme influençant les mesures (si faites sur le modèle) –** il existe de nombreux formalismes permettant de modéliser un processus. De plus, une modélisation reflète souvent un point de vue, impliquant des choix de modélisation. Dans le cas où les mesures portent sur l'évaluation du processus modélisé, le choix du formalisme utilisé pour la modélisation n'est pas anodin et peut induire un biais dans les résultats des mesures ;
- **évaluation non exhaustive –** certains aspects de la qualité d'un processus peuvent ne pas pouvoir être mesurés, soit parce que leur évaluation est trop coûteuse, soit parce que les informations nécessaires à l'évaluation sont indisponibles. Ceci rend l'évaluation non exhaustive et complexifie l'interprétation des résultats qui doit se faire en tenant compte d'informations manquantes ;
- **évaluation coûteuse –** certains aspects de la qualité d'un processus peuvent être coûteux à évaluer, par exemple s'ils nécessitent de mener une étude de terrain en menant un grand nombre d'interviews (ce qui peut être le cas lors de l'évaluation d'un processus métier par exemple) ;
- **dimensions qualité interdépendantes –** les dimensions qualité sont interdépendantes. Un exemple classique concerne les dimensions de *complétude* et de *lisibilité*. Plus on complète un processus, plus (en règle générale) on le complexifie, ce qui amène souvent à dégrader sa lisibilité. Ces interdépendances posent problème pour l'interprétation des résultats (il faut pouvoir comprendre quelles mesures sont liées et pourquoi) et pour le choix des actions d'amélioration à entreprendre (une action d'amélioration destinée à améliorer une mesure ne doit pas amener à en dégrader une autre).

3. L'évaluation des artefacts

3.1. *Travaux existants*

La recherche sur l'évaluation des artefacts d'un SI se concentre sur l'évaluation des modèles produits pendant le développement. Par exemple, les modèles métier ont été étudiés dans (Aguilar, 2006, Rittgen, 2010), ceux des besoins dans (Kanjilal *et al.*, 2009), les modèles architecturaux dans (Nowak *et al.*, 2011) et ceux d'interface dans (Panach *et al.*, 2008). En IDM (ingénierie dirigée par les modèles), la qualité des modèles et méta-modèles se base surtout sur l'utilisation de mesures pour les modèles à objets (Ma *et al.*, 2004). Dans ce contexte, Solheim et Neple (2006) soulignent deux critères importants de qualité : la transformabilité et la modificabilité. Cependant, d'autres critères doivent être aussi examinés. Dans cette direction, Monperrus *et al.* (2008) ont proposé une mesure générique sur les méta-modèles qui permet de définir d'autres mesures afin d'évaluer des caractéristiques de qualité comme la fonctionnalité, fiabilité et maintenabilité. De manière plus globale, Le Maitre (2008) synthétise différents types de frameworks pour la qualité qui ont été proposés pour des artefacts tels que ceux décrits ci-dessous :

- Des frameworks hiérarchiques qui proposent une vue en arbre de la qualité. Le concept racine est la qualité globale qui est ensuite divisée en différents aspects pouvant eux-mêmes être divisés à nouveau. Les feuilles sont généralement associées aux mesures. L'un des exemples est le standard ISO9126 (ISO, 2001). Nous pouvons aussi trouver des travaux pour évaluer spécifiquement certains types d'artefacts comme, par exemple, Fernandez *et al.* (2009) qui propose un framework pour évaluer des modèles d'interfaces pour systèmes WEB.
- Des frameworks de formalisation qui proposent une méthodologie pour construire des mesures de qualité en utilisant les propriétés mathématiques que les mesures doivent satisfaire. Dans (Briand *et al.*, 1996) les auteurs présentent un ensemble de propriétés mathématiques que des mesures doivent satisfaire pour assurer leur validité. On peut aussi citer le travail de (Poels et Dedene, 1999) qui propose un framework fondé sur la théorie de la mesure pour définir des mesures pour l'évaluation de la qualité.
- Des frameworks de causalité qui mettent l'accent sur l'influence de certaines propriétés sur d'autres et essaient de mesurer la corrélation des propriétés. Pour cette catégorie, nous pouvons citer les travaux de (Kesh, 1995) et (Maes et Poels, 2006). Kesh a proposé un modèle, une méthodologie et des mesures pour évaluer la qualité des modèles Entité- Relation.
- Des frameworks sémiotiques basés sur la théorie (Lindland, 1994) conçus pour évaluer n'importe quel type de modèles. Dans (Lindland, 1994), la qualité est détaillée en syntaxique, sémantique et pragmatique. Son but est la compréhension. Même si (Moody *et al.*, 2003) a montré l'intérêt du framework de Lindland, ce dernier a été étendu pour ajouter dans l'aspect pragmatique la qualité organisationnelle et la qualité technique (Krogstie, 1998) ou pour prendre en compte les spécificités de l'IDM (Solheim, 2006). Nous avons utilisé le framework de Krogstie pour structurer son évaluation du nouveau langage de modélisation des processus métier (Cortes *et al.*, 2011).

Ces frameworks sont intéressants car ils définissent la qualité et structurent le raisonnement. Néanmoins quand ils n'aboutissent pas à des méthodes d'évaluation telles que des mesures, leur utilisabilité est réduite. Or l'évaluation dépend de l'artefact qu'on évalue, des caractéristiques qu'on est en train de considérer (fonctionnalité, utilisabilité, etc.) et de qui évalue (ingénieur, utilisateur, etc.).

Aussi indépendamment des frameworks, de nombreux travaux ont abordé la qualité des modèles et des langages par des évaluations spécifiques. Par exemple, nous avons exploité la définition de mesures pour les modèles de tâches (Oliveira, 2011). Nous avons aussi défini des mesures pour évaluer des diagrammes de flux de données « *data flow diagrams* » et réalisé des évaluations pratiques des diagrammes dans le cadre de grands systèmes d'information de l'industrie (Ramos *et al.*, 2004). De manière similaire, (Céret, 2010) a collecté des mesures et les a interprétées sur du code plutôt que sur des modèles. Il propose de mesurer la qualité du code produit à l'issue d'un processus de développement et d'en déduire des éléments de qualité des modèles dont est issu le code. Cette technique permet d'éviter la complexité de

définition de mesures sur les modèles, mais elle fait l'hypothèse forte de l'inférence inverse c'est-à-dire que certains éléments du code (ici modularité et réutilisabilité) sont le reflet des choix de modélisation. Dans (Cherfi *et al.*, 2002) des mesures pour l'évaluation de la qualité des modèles Entité-relation et UML ont été proposées. Ce travail insiste surtout sur le fait que la vision de la qualité est relative et qu'il faut tenir compte de l'auditoire lors de son évaluation. En effet il serait utile d'appuyer l'évaluation de la qualité par une assistance à l'amélioration qui est un point directement lié à l'interprétation des résultats. Dans (Mehmood *et al.*, 2011) une contribution dans ce sens est proposée.

Les évaluations expérimentales sont difficiles à réaliser. Par exemple, (Dupuy-Chessa, 2011) propose une étude de cas de l'utilisation d'un processus de conception. Si cette expérience a permis de faire évoluer le processus grâce à des commentaires qualitatifs, il a par contre été difficile d'obtenir des résultats plus quantitatifs : le profil des participants (des experts de différents domaines de l'informatique), la durée (8 jours) et la nature des tâches à réaliser (des modèles du domaine de l'interaction homme-machine et des modèles UML) limitent les possibilités de réelles validations. D'ailleurs (Moody *et al.*, 2003) souligne la limite des expérimentations menés avec des étudiants et leur capacité à valider des approches proposées. Paloudi (1999) a souligné la limitation des catégorisations existantes des parties prenantes insistant sur le fait que leur classification en utilisateurs finaux, informaticiens et professionnels est insuffisante. Dans (Mehmood *et al.*, 2011 ; Akoka *et al.*, 2009) les évaluations menées avec des professionnels montrent la difficulté à motiver et à impliquer les professionnels.

3.2. *Limites et difficultés de l'évaluation des artefacts*

Comme le montrent la diversité des travaux existant et leur manque d'homogénéité, l'évaluation de modèles de spécification est une tâche difficile. Les difficultés liées à assurer cette qualité sont tout autant liées à la collecte des données pour l'évaluation qu'à l'interprétation des résultats. Les limites et difficultés peuvent être résumées comme présenté ci-dessous :

- **les protocoles expérimentaux sont généralement complexes et difficile à définir -** L'évaluation peut être réalisée à l'aide d'expérimentations ou grâce à des mesures. Dans le cas des expérimentations, les protocoles expérimentaux sont généralement complexes et difficile à définir : combien faut-il de participants ? Quel doit-être leur profil ? Quelles tâches leur faire réaliser pour permettre une évaluation ? Comment limiter les biais ? etc. ;
- **la non disponibilité des professionnels** - Une fois le protocole établi, il reste à collecter les mesures et à jongler avec la disponibilité des participants, ainsi qu'avec les éventuels coûts de réalisation. L'expérimentation, pour jouer son rôle de validation doit supprimer tout biais. Cependant, la non disponibilité des professionnels et la difficulté à les impliquer et à avoir leurs retours expliquent

les expérimentations « laboratoire » faites avec des étudiants dont les formateurs sont ceux qui ont proposé les approches à valider et montrent leurs limitations ;

- **difficulté d'interpréter les mesures** - Une fois les mesures collectées, il reste à interpréter les mesures obtenues. Dans le cas d'évaluations expérimentales, un expert est nécessaire pour l'interprétation. Dans le cas des mesures, l'évaluation est simplifiée/automatisée, mais l'interprétation des résultats reste compliquée : il faut que les mesures s'accompagnent de seuils qui définissent la qualité des potentiels résultats. Ces seuils même s'ils sont validés, ne sont généralement pas suffisants et un expert reste primordial pour ne pas fournir une mauvaise interprétation des résultats. En effet, le contexte (taille du projet, caractéristiques de l'équipe de développement, caractéristique du deadline du projet, etc..) est déterminant mais est peu pris en compte.

4. L'évaluation des interfaces homme-machine

4.1. *Travaux existants*

L'évaluation des interfaces homme-machine est un domaine à part entière, ayant fait l'objet de nombreux travaux depuis environ une quarantaine d'années. Elle fait partie intégrante des processus centrés utilisateur ISO 9241-210 (2010). Pendant de nombreuses années, l'évaluation s'est basée sur deux critères principaux (Nielsen, 1993) : l'utilité et l'utilisabilité. A ce sujet, le critère d'utilisabilité a fait l'objet de la norme ISO 9241-11 (1998) ; il est défini comme « le degré selon lequel un produit peut être utilisé, par des utilisateurs identifiés, pour atteindre des buts définis avec efficacité, efficience et satisfaction, dans un contexte d'utilisation spécifié ». L'accessibilité logicielle a pris ensuite énormément d'importance, et fait désormais l'objet d'une norme à part entière : ISO 9241-171 (2008). De nouveaux critères liés à la persuasion interactive pourraient maintenant venir en complément des critères classiques, comme suggéré par (Nemery, 2012) en lien avec le domaine du décisionnel.

De nombreuses publications sont consacrées à l'évaluation des interfaces homme-machine. (p.e., Kolski (1997), Bastien et Scapin (2001) ou Sears et Jacko (2007)). Dans (Scapin et Bastien, 1997) a été proposé un ensemble de critères et recommandations pour l'évaluation de la qualité des interfaces homme-machine, qui sont largement utilisés par de nombreux chercheurs et industriels. Une analyse d'aspects méthodologiques provenant d'une large application de tests d'utilisabilité est proposée dans (Bastien, 2010). Dans (Lima *et al.*, 2009), des évaluations de la qualité des interface homme-machine de sites gouvernemental exploitées par des utilisateurs souffrant de déficience visuelle sont décrites. Plusieurs travaux récents ont porté sur la mise en œuvre de nouveaux environnements dédiés à l'évaluation de systèmes interactifs, particulièrement à ceux possédant une architecture à base d'agents d'interface (Ezzedine *et al.*, 2006 ; Tran *et al.*, 2008) ; d'autres ont visé à prendre en compte la diversité des versions et évolutions de la méthode d'inspection *cognitive walkthrough* (Mahatody *et al.*, 2010).

4.2. *Limites et difficultés de l'évaluation des interfaces homme-machine*

L'évaluation de la partie interactive des systèmes d'information est cruciale dans la mesure où les systèmes d'information sont de plus en plus présents dans tous les domaines ; ils sont amenés à être utilisés par des utilisateurs de profils très variés. Cependant, à partir de nos connaissances de la littérature et de nos expériences de recherche présentées dans la section précédente, différentes difficultés et limites relatives à l'évaluation de la partie interactive des systèmes d'information sont soulignées ci-dessous :

- **prise en compte insuffisante des caractéristiques des organisations** – alors que l'aspect organisationnel est très important dans les systèmes d'information, la plupart des méthodes d'évaluation des IHM s'adressent à des situations mettant en œuvre des utilisateurs, sans particulièrement les situer par rapport à une organisation ; les aspects collectifs et collaboratifs des situations à considérer en évaluation des IHM mériteraient d'être explicitement prises en compte dans différentes catégories de méthodes ;
- **difficulté d'évaluer la partie interactive des systèmes d'information adaptatifs** – l'évaluation des interfaces homme-machine et systèmes interactifs dits adaptatifs (certains étant dits aussi intelligents ou sensibles au contexte, voire personnalisés) est un problème difficile, pour lequel peu de méthodes, outils et critères ont été proposés, ou en tout cas font actuellement l'unanimité ;
- **inexistence des modèles et profils utilisateur unifiés pour l'évaluation des SI** – malgré une importance notable apportée à la modélisation de l'utilisateur depuis une trentaine d'années en interaction homme-machine, une question importante se pose dans le domaine de l'évaluation des systèmes d'information : comment modéliser précisément différentes catégories ou profils d'utilisateurs des SI ? Il n'existe en effet pas encore de modèle unifié ;
- **insuffisance d'outils d'évaluation adaptés à l'évaluation des SI** – un problème récurrent en évaluation des systèmes interactifs réside dans l'insuffisance d'outils pour l'évaluation, combinant si possible un ensemble de méthodes à ce sujet.

5. L'évaluation des données

5.1. *Travaux Existants*

La qualité des données est un concept complexe et multidimensionnel (Redman, 1996) recouvrant différentes sémantiques selon le contexte. Elle est caractérisée selon un ensemble de dimensions, chaque dimension regroupant un ensemble de mesures. Nous avons décomposé les travaux sur la qualité des données en trois catégories : la définition des dimensions de la qualité, les techniques et approches pour la qualité des données et les outils pour la qualité des données.

Lorsqu'on parle de dimensions de la qualité des données on entend par là des concepts tels que l'exactitude (*accuracy*) qualifiant la part de données dont la proximité entre la valeur de la donnée et la valeur réelle est acceptable, la complétude qualifiant la part d'information manquante dans les données, la cohérence qualifiant la part de données respectant un ensemble de contraintes ou règles métier, ou encore l'actualité qualifiant la part de données obsolètes. De nombreuses autres dimensions ont été définies telles que l'accessibilité, la précision, la compréhensibilité, etc. (Wang *et al.*, 1995). Un grand nombre de classifications de ces dimensions fut d'ailleurs proposé dans la littérature (Batini et Scannapieca, 2006). Les dimensions sont interdépendantes, c'est-à-dire que les dimensions s'influencent les unes les autres (voir par exemple (Barone *et al.*, 2010)).

Il existe trois approches pour définir les dimensions pour la qualité des données. Les approches théoriques, les approches empiriques et les approches intuitives. Parmi les approches théoriques, Wand et Wang. (1996) ont représenté les concepts de « monde réel » et de « système d'information ». Ils ont ensuite rapporté le problème de qualité des données à une mise en correspondance entre ces deux mondes et ont formulé les écarts comme des « défauts » de qualité à partir desquels ils ont dérivé des dimensions de qualité telles que l'exactitude, la complétude ou la cohérence. Les approches empiriques utilisent les expérimentations. Dans (Wang et Strong, 1996) les dimensions de la qualité ont été définies en interrogeant des utilisateurs de données. Ils ont ainsi dérivé quatre catégories (qualité intrinsèque, contextuelle, représentation et d'accessibilité). Les approches intuitives utilisent une catégorisation choisie intuitivement. Dans (Redman, 1996) la qualité des données est classée selon la nature, la valeur et le format.

Concernant les méthodes et techniques, un grand nombre de méthodologies pour analyser et améliorer la qualité des données ont été proposées. Batini *et al.*, (2009) en proposent une revue et une comparaison. Des étapes communes classiques se dégagent de ces méthodologies (De Feo et Barnard, 2005)) : (1) la *définition* de la qualité des données, (2) la *mesure* de chacune des mesures, (3) l'*analyse* des résultats de l'évaluation, (4) l'*amélioration* de la qualité, et (5) le *contrôle* des effets de l'amélioration. Les techniques utilisées pour chacune de ces phases varient en fonction du type des données considérées (structurées, semi-structurées, non structurées), du type de système d'information gérant les données (par exemple, entrepôt, distribué, pair à pair), des dimensions qualité à considérer et des stratégies de mesures ou amélioration.

Enfin concernant les outils, une grande diversité d'outils s'adresse à la qualité des données. Tout d'abord nous trouvons les outils d'ETL (*Extract, Transform-Load*) qui intègrent une étape de nettoyage des données. Il existe une grande offre commerciale pour les outils de standardisation d'adresses. Une étude comparative est présentée dans (Barateiro et Galhardas . 2005).

5.2. *Limites et difficultés de l'évaluation des données*

Nous nous focalisons ici sur l'évaluation de la qualité des données que nous considérons composée des étapes *définition*, *mesure* et *analyse*. Notre expérience nous a amenés à exhiber quelques difficultés et limites d'une évaluation, que nous décrivons ci-dessous :

- **grand nombre de mesures disponibles pour évaluer la qualité des données** – les méthodologies pour la définition de mesures ne fournit qu'un cadre général et le choix des mesures, parmi le très grand nombre de mesures disponibles, reste toujours une activité délicate ;
- **difficulté à associer un seuil à une mesure** – il est courant de vouloir associer un (ou plusieurs) seuil(s) permettant d'associer un niveau d'alerte à une mesure. Cependant, il est difficile de fixer les valeurs de ces seuils. L'utilité et surtout l'applicabilité de ces seuils reste discutables puisque généralement cette valeur dépend de plusieurs paramètres dont le contexte d'évaluation lui-même ;
- **résultats incomplets ou approximatifs** – il n'est pas toujours possible de mesurer de façon exacte une dimension/mesure. Il existe plusieurs raisons à cela. (i) Soit la dimension/mesure en elle-même ne permet pas un calcul exact. Nous prenons par exemple ici l'exemple de mesures associées à la dimension d'*exactitude* pour lesquelles l'exactitude est souvent approchée (par le calcul de l'exactitude syntaxique, par vérification de cohérence, etc.) (ii) Soit parce que la mesure nécessite une mesure subjective. Il s'agit par exemple ici des mesures obtenues par interviews. (iii) Soit pour des problèmes plutôt « techniques » : une partie des données n'est pas disponible ou les données sont de trop gros volume. Dans ce cas, seul un échantillon des données est étudié pour calculer la mesure, les résultats sont donc incomplets ou approximatifs. Il se pose d'ailleurs dans ce cas une autre difficulté qui est le choix de l'échantillon et de la technique de redressement à adopter ;
- **L'interprétation des résultats est toujours compliquée** même pour des experts métier. Les influences entre mesures sont importantes à identifier car dans le cas de mesures s'influençant « négativement », une action d'amélioration destinée à améliorer une mesure pourrait amener à dégrader une autre mesure. Berti-Equille *et al.*(2011) discutent ce problème et proposent une approche permettant de l'étudier.

6. Synthèse

Comme expliqué ci-avant, l'analyse présentée dans les sections précédentes est réalisée en se basant sur les travaux existants sur l'évaluation de chaque « dimension » (processus, artefacts, interfaces et données) individuellement. Nous étions conscients du fait que nous pourrions identifier des limites similaires sur l'analyse des différentes dimensions, mais nous avons choisi de les examiner séparément pour pouvoir ensuite les confronter et réaliser la synthèse présentée dans

cette section. Le Tableau 1 présente le résultat de cette synthèse, où ✓ indique où chaque limite a été identifiée et ☑ où nous considérons que la limite est aussi pertinente après une analyse en confrontant les difficultés avec les autres dimensions. Nous remarquons que certaines limites sont pertinentes pour toutes les dimensions (comme par exemple, le grand nombre de mesures et la difficulté de collecter et interpréter ces mesures), et que d'autres sont spécifiques à une dimension (comme p. e. la modélisation et les profils utilisateur pour l'évaluation des SI qui influence directement l'évaluation des interfaces homme-machine et ne concerne que cette dimension).

Limites et difficultés	**P***	**A**	**I**	**D**
1. Grand nombre de mesures	✓	☑	☑	✓
2. Difficulté de collecter et interpréter les mesures	✓	✓	☑	✓
3. Difficulté à associer un seuil à une mesure et l'utiliser	✓	☑	☑	✓
4. Difficulté d'évaluer à l'avance si la modification doit être faite	✓			
5. Choix de modélisation et choix du formalisme influençant les mesures (si faites sur le modèle)	✓	☑		
6. Évaluation non exhaustive	✓		☑	
7. Évaluation coûteuse	✓	☑	☑	☑
8. Dimensions qualité interdépendantes	✓	☑	☑	☑
9. Protocoles expérimentaux sont généralement complexes et difficiles à définir	☑	✓	☑	☑
10. Non disponibilité des professionnels pour réaliser les expérimentations	☑	✓	☑	☑
11. Prise en compte insuffisante des caractéristiques des organisations			✓	
12. Difficulté d'évaluer la partie interactive des SI adaptatifs			✓	
13. Inexistence des modèles et profils utilisateur unifiés pour l'évaluation des SI			✓	
14. Insuffisance d'outils d'évaluation adaptés à l'évaluation des SI	☑	☑	✓	☑

*P-Processus, A-Artefacts, I-Interfaces, D-Données

Tableau 1. *Limites et difficultés de l'évaluation d'un SI*

7. Conclusion et perspectives

Le tour d'horizon présenté ici montre tout d'abord la prise de conscience quant à l'importance de l'évaluation de la qualité au plus tôt. La prolifération des propositions a très vite mis l'accent sur la nécessité de leur validation. Ce travail a montré un certain nombre de manques et par conséquent de perspectives de recherche, que nous pouvons généraliser comme suit :

– malgré la richesse des propositions faites par la communauté des chercheurs, ces approches sont peu connues et peu utilisées par les professionnels de l'informatique. Pour améliorer cette situation quelques idées doivent être développées, telles que le développement d'outils qui implémentent ces

approches ou l'implication des professionnels dans les manifestations scientifiques (ateliers, sessions pour industriels etc.) ;

- la difficulté à mener des expérimentations avec les professionnels et la difficulté à les motiver nous indiquent le besoin d'exploiter l'utilisation des nouvelles technologies ou outils pour attirer la participation des professionnels dans la réalisation des évaluations ;
- Peu de travaux valorisent la validation théorique alors que compte tenu des limites des validations empiriques, centrées sur l'expérimentation, cette voie paraît indispensable à la validation des approches de qualité ;
- Enfin, la prolifération des propositions et le manque de fédération fait ressortir le besoin de mener un travail de fédération et de conciliation des travaux existants.

Nous travaillons maintenant vers la définition d'actions de recherche et de pratiques pour répondre aux difficultés identifiées en espérant trouver des solutions concrètes pour l'évaluation des SI.

Remerciements

Les auteurs remercient l'association INFORSID pour le financement à l'action « Évaluation des SI » sans lequel ce travail ne pourrait pas avoir été réalisé.

Bibliographie

Aguilar, E. R., Ruiz, F., Garcia, F., Piatinni, M., « Evaluation Measures for Business Process Models », *Proc. of the ACM Symp. on Applied Computing* (*SAC'06)*, 2006, p. 1567-1568.

Akoka J, Comyn-Wattiau I., Cherfi S.S.-S., « Qualité perçue des schémas conceptuels. Etude comparée informaticiens vs. utilisateurs », *Ingénierie des Systèmes d'Information,* vol. 14, n°4, 2009, p. 33-54.

Almeida M.A., Blanc X., Bendraou R., « Deviation Management during Process Execution », *Proc. of Intl. Conf. on Automated Software Engineering*, 2011, p. 528-531.

Almeida M.A., Mougenot A., Bendraou R., Robin J., Blanc X., « Artifact or Process Guidance, an Empirical Study », *Proc. of MODELS 2010*, LNCS 6395, 2010, p. 318-330.

Barateiro J., Galhardas H., « A Survey of Data Quality Tools », *Datenbank-Spektrum*, vol. 14, 2005, p. 15-21.

Barone D., Stella F., Batini C., « Dependency Discovery in Data Quality », *Proc. of the Intl. Conf. in Advanced Information Systems Engineering (CAiSE)*, 2010, p. 53-67.

Basili V.R., Caldiera G., Rombach H.D., « The Goal Question Metric Approach », *Encyclopedia of Software Engineering*, Wiley, 1994.

Bastien C., Scapin D., « Évaluation des systèmes d'information et Critères Ergonomiques », *in Kolski C. (dir.), Environnements évolués et évaluation de l'IHM, Interaction homme-machine pour les SI 2*, Hermes, Paris, 2001, p. 53-80.

Bastien J. M. C., « Usability testing: a review of some methodological and technical aspects of the method », *Intl. Journal of Medical Informatics*, vol. 79, n°4, 2010, p.18-23.

Batini C., Cappiello C., Francalanci C., Maurino A., « Methodologies for data quality assessment and improvement », *ACM Comput. Surv.*, vol. 41, n°3, 2009, p. 1-52.

Batini C., Scannapieca M., *Data quality: concepts, methodologies and techniques*, Springer, 2006.

Berti-Equille L., Comyn-Wattiau I., Cosquer M., Kedad Z., Nugier S., Peralta V., Si-Said Cherfi S., Thion-Goasdoué V.: « Assessment and analysis of information quality: a multidimensional model and case studies ». *Intl. Journal of Information Quality*, vol. 2, n°4, 2011, p. 300-323.

Briand L., Morasca S., Basili V., « Property-Based Software Engineering Measurement », *IEEE Transactions on Software Engineering*, vol. 22, n°1, 1996, p. 68–86.

Ceret E., Dupuy-Chessa S., Godet-Bar G., « Using Software Metrics in the Evaluation of a Conceptual Component », *Proc. of the 4th RCIS'2010*, 2010, Nice, France, p. 507-514.

Cherfi S. Si-Said, Akoka J., and Comyn-Wattiau I., « Conceptual Modeling Quality - From EER to UML Schemas Evaluation ». *Proc. of the 21st Intl. Conf. on Conceptual Modeling (ER '02)*, 2002, p. 414-428.

Chrissis, M., Konrad M., Shrum S. *CMMI: guidelines for process integration and product improvement*. 2.ed. Boston: Addison-Wesley, 2003.

Comyn-Wattiau I., Akoka J., Berti-Equille L., « La qualité des systèmes d'information : Vers une vision plus intégrée », *Revue des Sciences et Technologies de l'Information série ISI*, vol. 15, n° 6, 2010, p. 9-32.

Cortes-Cornax M., Dupuy-Chessa S., Rieu D., and Dumas M., « Evaluating choreographies in BPMN 2.0 using an extended quality framework », *Proc. of the 3rd Intl. Workshop on the Business Process Model and Notation*, 2011.

De Feo J. A., Barnard W. *JURAN Institute's Six Sigma Breakthrough and Beyond - Quality Performance Breakthrough Methods*. McGraw-Hill Professional, 2005.

Dupuy-Chessa S., Mandran N., Godet-Bar G. and Rieu D., « A case Study for Improving a Collaborative Design Process », *Proc of IFIP WG8.1 Working Conf. on Method Engineering (ME'2011)*, Paris, France, April 2011.

Ezzedine H., Trabelsi A., Kolski C., « Modelling of an interactive system with an agent-based architecture using Petri nets, application of the method to the supervision of a transport system », *Mathematics and Computers in Simulation*, vol. 70, 2006, p. 358-376.

Fernandez A., Insfran E. and Abrahão S., « Integrating a Usability Model into Model-Driven Web Development Processes », *Proc. of WISE'2009, LNCS 5802*, p. 497–510, 2009.

Grim-Yefsah M., Rosenthal-Sabroux C., Thion-Goasdoué V., « Évaluation de la qualité d'un processus métier à l'aide d'informations issues de réseaux informels », *Revue des Sciences et Technologies de l'Information série ISI*, vol 15, n° 6, 2010, p. 63-83.

Horie D., Kasahara T., Goto Y., Cheng J., « A New Model of Software Life Cycle Processes for Consistent Design, Development, Management, and Maintenance of Secure Information Systems », *Proc. of ICIS*, 2009, p. 897-902.

Hwang S.M., « Process Quality Levels of ISO/IEC 15504, CMMI and K-model », *Intl. Journal of Software Engineering and its Applications*, vol. 3, n°1, 2009.

Irani Z., « Information systems evaluation: navigating through the problem domain », *Information and management*, vol. 30, 2002, p.11-24.

ISO 9000:2000, Quality management systems — Fundamentals and vocabulary, 2000.

ISO 9241-11, Ergonomic requirements for office work with visual display terminals, 1998.

ISO 9241-171, Ergonomics of human-system interaction — Part 171: Guidance on software accessibility, 2008.

ISO 99241-210, Ergonomics of human-system interaction — Part 210: Human-centred design for interactive systems, 2010.

ISO/IEC 12207 Systems and Software Engineering - Software Life Cycle Processes, 2008.

ISO/IEC 15504-3 Information technology — Process assessment — Part 3: Guidance on performing an assessment, 2004.

ISO/IEC 9126, Software engineering - Product quality - Part 1: Quality model, 2001.

Kanjilal A., Sengupta S., Bhattacharya S., « Analysis of Complexity of Requirements: A Metrics based Approach », *Proc. of ISEC*, 2009, p. 131-132.

Kesh S., « Evaluating the quality of entity relationship models ». *Information and Software Technology*, 1995, p. 681-689.

Kolski C., *Interfaces homme-machine*, Paris, Hermès, 1997.

Krogstie J., "Integrating the Understanding of Quality in Requirements Specification and Conceptual Modeling", Software Engineering Notes, ACM SIGSOFT, 23(1), pp 86-91, Janvier 1998.

Lanubile F., Visaggio G., « Evaluating defect detection techniques for software requirements inspections », ISERN-00-08, Intl Software Engineering Research Network, 2000.

Lemaitre J. et Hainaut J.L., « A combined global-analytical quality framework for data models, workshop on quality in modelling at models », Computer Science Faculty, FUNDP, Namur, Belgium, 2008.

Lima S., Lima F., Oliveira K. M., « Evaluating the Accessibility of Websites to Define Indicators in Service Level Agreements », *Proc. of the 11th Intl. Conf. on Enterprise Information Systems*, 2009.

Lindland O.I., Guttorm Sindre et Arne Sølvberg, « Understanding quality in conceptual modeling », *IEEE Software*, p. 11 :42-49, 1994.

Ma H., Shao W., Zhang L., Ma Z., and Jiang Y., « Applying OO metrics to assess UML metamodels », *Proc. of MODELS/UML*, LNCS 3273, 2004, p. 12-26.

Machado L. F., Oliveira K. M., Rocha A.R., « Using Standards And Maturity Models for the Software Process Definition », *Proc. of the 4th Intl Software Quality Week Europe*, 2000.

Maes A., Poels G., Evaluating quality of conceptual models based on user perceptions. *Proc of the Intl.Conf. on Conceptual Modeling*, 2006, p. 54-67.

Mahatody, T. Sagar, M. Kolski, C., « State of the Art on the Cognitive Walkthrough method », its variants and evolutions. *Intl. Journal of Human-Computer Interaction*, vol. 26, 2010, p. 741-785.

Mehmood, K.,Cherfi, S.S.-S.,Comyn-Wattiau, I.,Akoka, J., « A pattern-oriented methodology for conceptual modeling evaluation and improvement », *Proc. of 5th RCIS*, 2011, p. 1-11.

Monperrus M., Jezequel J-M, Champeau J. and Hoeltzener B., « Model-Driven Software Development: Integrating Quality Assurance », *Measuring Models. — IDEA Group Inc.*, Hershey, PA, USA, 2008.

Monteiro L.F.S., Oliveira K.M., « Defining a catalog of indicators to support process performance analysis », *Journal of Software Maintenance,* vol. 23, n°6, 2011, p. 395-422.

Moody D., Sindre G., Brasethvik T., Solvberg A., « Evaluating the Quality of Information Models : Empirical Testing of a Conceptual Model Quality Framework », *Proc. of the 25th Intl. Conf. on Software Engineering*, 2003, p. 295-305.

Nemery A., « Elaboration, validation et application de la grille de critères de persuasion interactive », Thèse de doctorat, Université de Lorraine, Janvier 2012.

Nielsen J., *Usability Engineering*. Academic Press, Boston, 1993.

Nowak M., Pautasso, C. « Goals, Questions and Metrics for Architectural Decision Models », *Proc. of SHARK*, 2011, p. 21-28.

Oliveira K., « Une étude initiale sur les mesures d'utilisabilité pour les modèles de tâches exprimés avec CTT », Présentation de *position paper*, INFORSID, 2011.

Paloudi, A., « Aspects of the Stakeholder Concept and their Implications for Information Systems Development ». *Proc. of the Thirty-second Annual Hawaii International Conference on System Sciences*, vol. 7. Washington, DC, USA, 1999.

Panach J., Condori-Fernández N., Valverde F., Aquino, N., Pastor, O., « Towards an Early Usability Evaluation for Web Applications », *IWSM-Mensura 2007*, LNCS 4895, p. 32-45, 2008.

Poels G., Dedene G., « DISTANCE: A Framework for Software Measure Construction », Research Report DTEW9937, Dept. Applied Economics Katholieke Universiteit Leuven, Belgium, 1999.

Raffo D., « Software project management using PROMPT: A hybrid metrics, modeling and utility framework », *Information and Software Technology*, vol. 47, n°15, 2005, p. 1009-1017.

Ramos C. S., Oliveira K. M., Anquetil N., « Legacy Software Evaluation Model for Outsourced Maintainer », *Proc. of the 8th European Conf. on Software Maintenance and Reengineering*, IEEE Computer Society, 2004, p. 48-57.

Redman T . C., « The impact of poor data quality on the typical enterprise», *Communications of the ACM*, vol. 41, n°2, 1998, p. 79-82.

Reix R., *Systèmes d'information et management des organisations*, 4e éd., Vuibert, 2002.

Rittgen, P. « Quality and Perceived Usefulness of Process Models », *Proc. of SAC'10*, 2010, p.65-72.

Rocha A. R., Montoni M., Santos G., Oliveira *et al.*, « Success Factors and Difficulties in Software Process Deployment Experiences based on CMMI and MR-MPS.BR », *Proc. of the Intl. Workshop on Learning Software Organizations* (LSO), 2006, p.77-87.

Santos R. P. dos, Oliveira K. M. de, Silva W. P. da., « Evaluating the Service Quality of Software Providers appraised in CMM/CMMI », *Software Quality Journal*, Springer Verlag, vol. 17, n°3, 2009, p. 283-301.

Scapin D. L., Bastien J. M. C. « Ergonomic criteria for evaluating the ergonomic quality of interactive systems », *Behaviour & Information Technology,* vol. 16, 1997, p. 220-231.

Sears A., Jacko J.A. (Eds.), *The Human-Computer Interaction Handbook: Fundamentals, Evolving Technologies and Emerging Applications* », Second edition. CRC Press, 2007.

SEI, Software Engineering Institute. *CMMI® for Development, Version 1.2*. Pittsburgh, PA, EUA 2011.

Solheim, I. and Neple, T. « Model Quality in the Context of Model-Driven Development », *Proc. of the 2nd Intl. Workshop on Model-Driven Enterprise Information Systems (MDEIS'06)*, 2006, p. 27-35.

Staples M., Niazi M., « Two case studies on small enterprise motivation and readiness for CMMI », *Proc. of the 11th Intl. Conf. on Product Focused Software*, 2010, p. 63-66.

Stylianou A. C., Kumar R. L., « An integrative framework for IS quality management », *Communications of the ACM*, vol. 43, n°9, 2000, p. 99-104.

Tran C-D., Ezzedine H., Kolski C., « Evaluation of agent-based interactive systems: proposal of an electronic informer using Petri Nets », *Journal of Universal Computer Science*, vol. 14, p. 3202-3216, 2008.

Wand Y., Wang R.Y., « Anchoring Data Quality Dimensions in Ontological Foundations », *Communications of the ACM*, vol. 39, n°11, 1996, p. 86-95.

Wang R. Y., Storey V. C., Firth C. P., « A Framework for Analysis of Data Quality Research », *IEEE Trans. on Knowl. and Data Eng.*, vol. 7, n°4, 1995, p. 623-640.

Wang R. Y., Strong D. M., « Beyond accuracy: what data quality means to data consumers », *Journal of Management Information Systems*, vol. 12, n°4, 1996, p. 5-33.