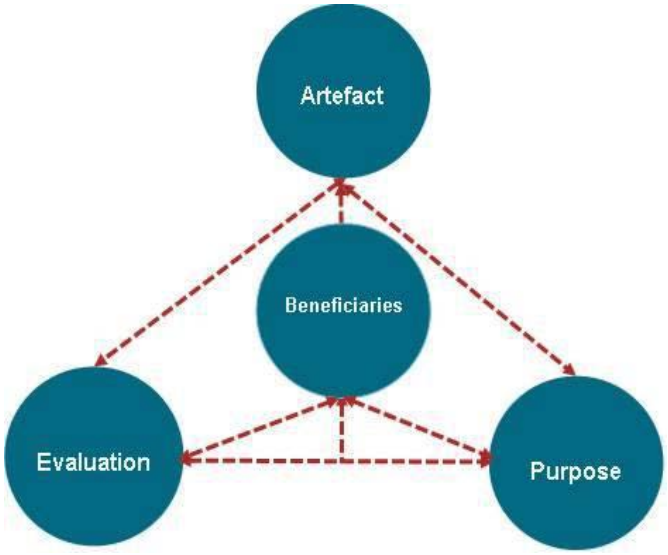
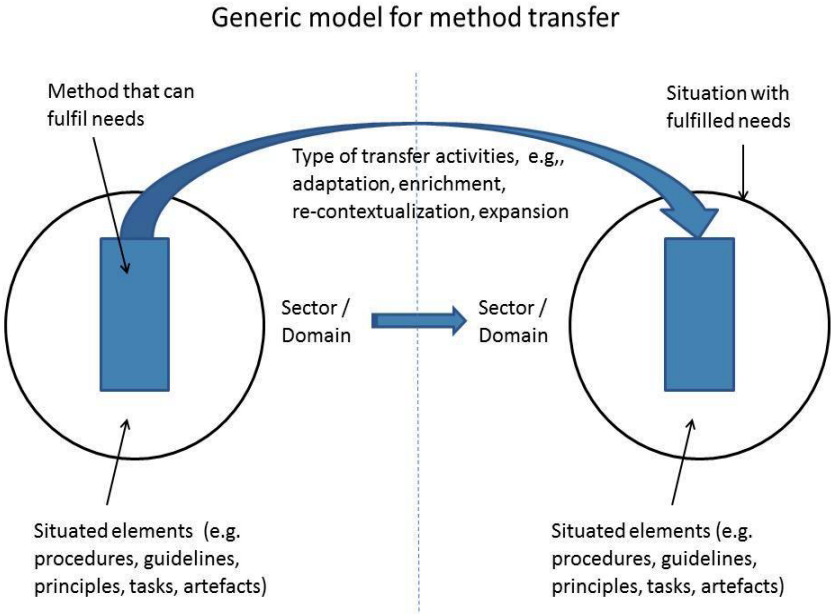
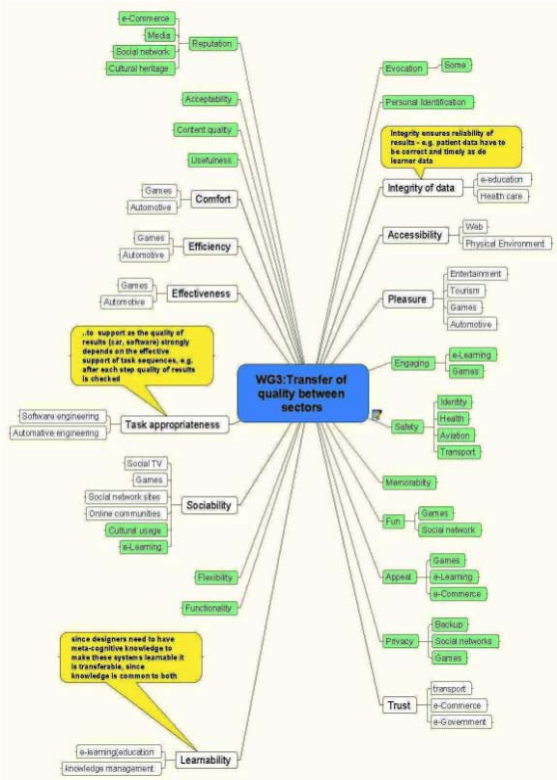
Towards the integration of transectorial IT design and evaluation

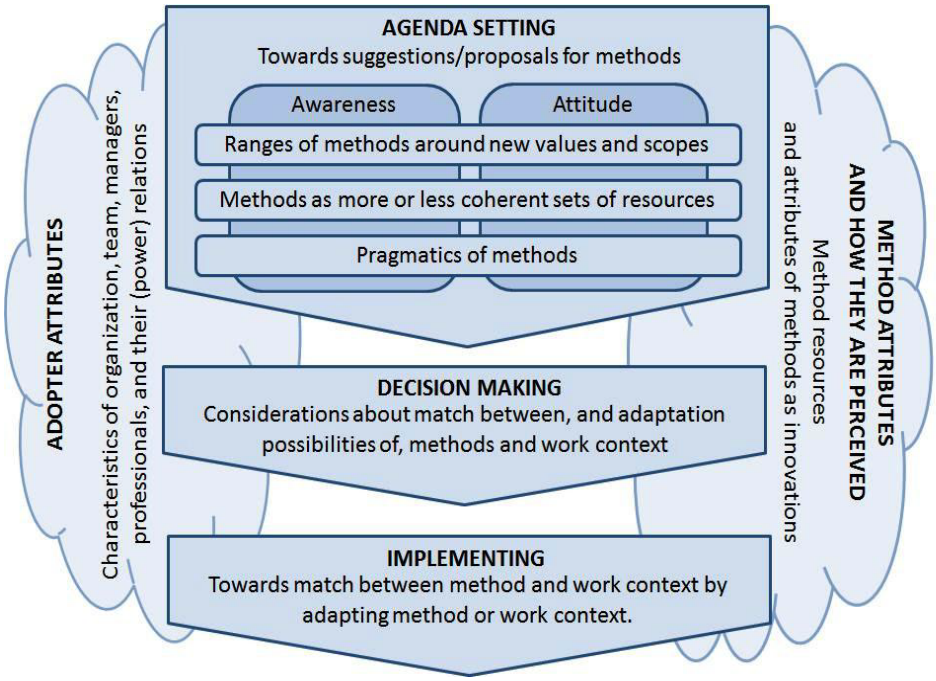
IC0904 TwinTide




# COST IC0904 TwinTide Working Groups Final Report




Editor: **Effie Law**

Author: **Gilbert Cockton, Christian Stary, Matthias Neubauer, Ebba Hvannberg, Arnold Vermeeren**

# Acknowledgements

On behalf of the consortium, I would like to express gratitude to our sponsor – COST (European Cooperation in the field of Scientific and Technical Research; http://cost.cordis.lu/src/home.cfm). The COST Office operated by the European Science Foundation (ESF) provides scientific, financial and administrative support to its Actions. Our Action IC0904 (http://www.twintide.org), which is variously known as **TwinTide (Towards the Integration of Trans-sectorial IT Design and Evaluation,** has been run in the domain of **Information and Communication Technologies (ICT)** for four years from 2009-2013. During those years, we have received continuous support from the members of the COST ICT Directorate, our Scientific Officer (*Dr. Ralph Stübner*), Administrative Office (*Ms. Aranzazu Sanchez*) and Rapporteur (*Prof. Juan Garbajosa)*, contributing to the success of the Action.

Certainly, the dedicated efforts of the Working Group (WG) leaders, Dissemination leaders and vice-Chair have been much needed and also highly appreciated. Their enthusiasms and passions have resonated with those of the Action's individual members. The network substantiated through the Action's activities has sowed many seeds for collaboration, which will continue to grow, blossom and come to fruition in the coming years.

*Effie Law*
*(Chair of COST IC0904 TwinTide)*

# Table of Contents

# COST IC0904 TwinTide

## Retrospect & Prospect

**Effie Lai-Chong Law**

University of Leicester, UK/
ETH Zürich, Switzerland

# COST IC0904 TwinTide: Retrospect and Prospect

Effie Lai-Chong Law

Department of Computer Science, University of Leicester, University Road, LE1 7RH, Leicester, UK/ ETH Zürich, TIK, Gloriastrasse 35, CH 8092, Zürich Switzerland
elaw@mcs.le.ac.uk/ law@tik.ee.ethz.ch

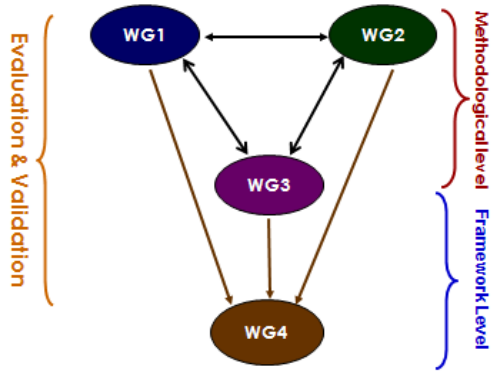
**Abstract.** An overview of COST Action IC0904 TwinTide (2009-2013) is presented in this chapter, outlining the basic structure and people behind this research network. Amongst others, some scientific achievements of TwinTide are highlighted, including re-conceptualization of three key notions – sector, transfer and method, identification of method transfer model and of building blocks for a generic method-selection-and-application framework, and training of young researchers. Outlook for the future development is sketched.

## 1 Overview of TwinTide

### 1.1 What is TwinTide?

COST (European Cooperation in the field of Scientific and Technical Research; http://cost.cordis.lu/src/home.cfm) operated by the European Science Foundation (ESF) provides scientific, financial and administrative support to its Actions of different domains. The COST Action IC0904 (http://www.twintide.org), which is variously known as TwinTide (*Towards the Integration of Trans-sectorial IT Design and Evaluation*), was officially launched on 26th November 2009 under the domain of Information and Communication Technologies (ICT) and concluded on 25th November 2013. The main goal of TwinTide was to integrate and foster research and practice on design and evaluation methodologies for computing artefacts across sectors and disciplines. TwinTide consisted of four working groups (WG), which were interdependent and at the same time had respective objectives geared towards the attainment of the Action's overarching goal (Figure 1).

Figure 1: Interdependencies of TwinTide Working Groups



During the first two years of the Action, WG1 ("Critical Review of Design & Evaluation Method Potential and Value-Realisation in IT-enhanced Sectors") and WG2 ("Transferability of Design and Evaluation Methods across IT-enhanced Sectors") laid the foundation work of the Action by systematically analysing the key notions, namely method, sector and transfer. In WG2, techniques such as concept maps, question café and case studies were employed to gain in-depth understanding of the transferability of methods. Built on this foundation work, during the third year of TwinTide, WG3 ("Interplay between Design and Evaluation, Quality Models and Standards") studied the quality attributes and use scenarios amenable to method transfer whereas WG4 ("Integrated D&E Methodological Framework") examined criteria for evaluating existing tools supporting method selection and application. In the fourth and also last year of the Action, integration and consolidation work was implemented to meet the Action's major challenge - constructing a generic method selection and application framework that effectively leverage and support interdisciplinary and cross-sectorial design and evaluation solutions.

### 1.2 Who is behind TwinTide?

Like most of today's collaborative projects, TwinTide could only be run effectively with the contributions of its partners, who came from 28 European countries. To facilitate the Action's progress, a steering committee comprising the following members was formed: Chair (*Effie L-C Law*, University of Leicester/ETH Zürich), Vice-chair (*Mark Springett*, Middlesex University, UK), four WG leaders (see Section 2) and Dissemination Leaders (*Marco Winckler & Philippe Palanque*, Université Toulouse III - Paul Sabatier, France). TwinTide was a large-scale consortium with more than 50 Management Committee (MC) members, who were involved in one or more WGs. In addition, about 15 non-MC members were involved in the activities of WGs. Their contributions to the success of the Action are indispensable, though it is impractical to name each of them individually here (see http://www.twintide.org/[partners]).

## 2 Scientific Achievements of TwinTide

The main body of this volume comprises four comprehensive reports of individual Working Groups (WGs). Here below we outline the key features of each WG, leaving the details to the corresponding reports.

**WG1:** Title*:* Critical Reviews on Design & Evaluation (D&E) Method Potential and Value-Realisation in IT-enhanced Sectors
Leader: *Gilbert Cockton*, Northumbria University, UK
Objective: WG1 aims to lay down the foundation of the Action by identifying and critically reviewing cases of D&E method applications
Task: (i) Prepare overviews of key sector exemplars for TwinTide; (ii) Detailed collection and analysis of sectors – building and prioritizing a list of

sectors, and establishing a common understanding of the boundary conditions for a sector; (iii) Collect a critical mass of analysis and D&E methods; (iv) Sustain a knowledge pool of relevant research methods; (v) Identify relevant technologies and their relationships to sectors.

**WG2:** Title: Transferability of D&E Methods across IT-enhanced Sectors
Leader: *Christian Stary*, University of Linz, Austria
Objective: WG2 aims to set the stage for developing a notion and an operational framework for transferability of D&E methods, which will be studied along several tasks in a collaborative design effort
Task: (i) Collection of documented attempts to transfer D&E methods across sectors, either being successful or having failed; (ii) Identification of criteria for assessing method transferability. This task tackles issues such as granularity of information and semiotic questions, in particular with respect to the expected outcome (value realization) when methods are transferred; (iii) Assessing (relative) importance of criteria to create a critical (optimal) set for assessing methods transferability between sectors.

**WG3:** Title: Interplay between Design and Evaluation, Quality Models and Standards
Leader: *Ebba Hvannberg*, University of Iceland
Objective: The main objective is of WG3 to identify and critically review the quality attributes and software standards that are currently prioritized in D&E processes in different sectors.
Task: (i) Identify software in use qualities. Examine how they vary over lifecycle and their dependency on sectors. How they are prioritized, traded off and realized in software processes; (ii) Identify which standards address emerging software qualities and what their impacts on real practice are; (iii) Understand iterative design-evaluation-redesign feedback cycles.

**WG4:** Title: Integrated D&E Methodological Framework
Leader: Arnold P.O.S. Vermeeren, TU Delft, the Netherlands
Objective: The aim is to develop an integrated D&E methods selection framework that should help practitioners in selecting methods for use in their specific situations. Additionally, the framework should serve as a reference model for future development of HCI in terms of training, education and identifying opportunities for developing new methods.
Task: (i) synthesizing a framework of D&E methods based on outcomes of the other three WGs; (ii) developing decision support mechanisms for selecting D&E methods.

The work of the four WGs resulted in a range of scientific achievements of which we highlight several in the following:

- **Re-conceptualisation of three key notions** – *sector*, *transfer* and *method* - can have significant impacts on a number of levels: research communities of various fields, industries of IT-enhanced sectors, standardisation bodies and individual users:

- In WG1, critical analyses and reviews of different classifications of sectors (e.g. SIC) have been conducted. Furthermore, the outcomes of WG1 tasks have led to insights how individual sectors shape and in turn are shaped by design and evaluation methods.
- Despite decades of research work (especially in the field of psychology), controversies on whether skills and knowledge are transferable across application contexts remain unsettled. Limited research work, however, has been done on the transfer of methods. In WG2 and WG3, the notion of transfer has been analysed systematically with the established methodologies.
- The notion of "HCI methods as sets of resources" is documented in Woolrych et al (2011)[1], which has contributed significantly to the work of WG1 and WG4. Based on this novel insight into the nature of HCI methods, a workshop entitled "Stories of Transfer, Triumph and Tragedy"[2] was held in CHI'13 (Paris, France). The main goal of the workshop was to further discuss this alternative conception about HCI methods with the larger HCI community and collect real-life case studies from practitioners to substantiate the preliminary version of the integrated method selection framework being constructed by WG4. The workshop attracted 18 quality papers.

- **Identification of Method Transfer Framework:** WG2 of the TwinTide aimed to identify and analyze sector-unique design and evaluation methodologies in terms of their underlying conceptual models, practical protocols, and transferability to other sectors. The results were intended to structure and foster the discussion about concepts and experiences when transferring methods across domains or sectors. As a result, method transfer is understood as the selection and application of a design or evaluation method in a specific development context that might have not been addressed so far when applying this method, but could qualify for the development situation at hand. Transparent method characteristics and context-of-use parameters referring to the development situation can facilitate the transfer of methodological knowledge. It is the reflected situation-sensitive match of parameters leading to informed design or evaluation decisions. The conceptual result and its application has been reported in the special issue of *Universal Access in the Information Society UAIS*) on 'Method Transfer across Domains and Disciplines: Enriching Universal-Access Development' (edited by C. Stary & S. Cronholm). The issue's intention is to structure and foster the discussion about concepts and experiences when transferring methods across domains or sectors. Situation dimensions become evident when considering the broad range of user characteristics, the changing nature of human activities, the variety of contexts of use, the increasing diversification of information, knowledge sources and e-services, and the proliferation of technology platforms. All invited authors focus on the increasing design challenges arising from the developments listed, and how

[1] Woolrych, A., Hornbæk, K., Frøkjær, E., and Cockton, G. (2011) Ingredients and meals rather than recipes: A proposal for research that does not treat usability evaluation methods as indivisible wholes. *International Journal of Human-Computer Interaction 27*(10):940-970.

[2] http://www2.le.ac.uk/departments/computer-science/people/elaw/HCI-3T

they can be handled by learning from specific cases or re-contextualizing methods supporting design or evaluation.

- **Building blocks for a generic method-selection-and application framework**: WG4 dealt with the challenge of integrating the insights from the other three WGs to construct this framework from the conceptual (e.g., definitional issue of method and transfer) and technological (e.g., tools analysis) perspective. Critical elements (or building blocks) for such a framework have been identified. While consolidating these elements into a coherent whole to address diverse constraints and conditions of different sectors remains to be explored in future work, some significant progress on this challenging research inquiry has been made through the development of *Diffusion of Design and Evaluation Method Innovation* (DIDEMI).
- **Training of young HCI researchers:** One of the main objectives of TwinTide was to train young researchers and provide them with opportunities to network with fellow as well as more experienced researchers, supporting the development of their future career. Three training schools for PhD students on research methods (DESIRE II[3]; TUTOREM[4]) were held. Altogether about 90 PhD students benefited from the schools. Apart from the training schools, Short-term Scientific Missions (STSMs) provided valuable opportunities for young researchers to visit their experienced counterpart to research on a specific topic over a period of time. A range of fruitful outcomes such as scientific publications, research proposals and career development has been produced.

## 3. Outlook

While TwinTide has made substantial progress in tackling the major challenge of developing a *generic method-selection-and-application* (GEMSA) framework, which can effectively leverage and support interdisciplinary and cross-sectorial design and evaluation solutions, more research efforts still need to be invested in meeting this challenge to a full extent. The complexity of the challenge lies in the fact that the methods are not static entities. This renders the task of mapping methods to contextual characteristics and constraints of individual application domains, which are inherently dynamic as well, particularly daunting. Furthermore, the repertoire of design and evaluation methods is expanding to address emerging technologies. Although some of such 'new' methods result from compiling and refining existing ones, the higher number of method options makes the selection even harder.

Clearly, the challenge of constructing a robust GEMSA framework entails further collaborative efforts. The research network enabled by TwinTide has paved not only a single route but a multitude of possible ones to meet this challenge. A follow-up COST Action as well as other opportunities to sustain the fruitful research network are being sought and much anticipated.

[3] http://www.tik.ee.ethz.ch/~lawl/DEVISEII/
[4] http://www2.le.ac.uk/departments/computer-science/people/elaw/tutorem/

# TwinTide WG1 Report

## Critical Reviews on Design and Evaluation Method Potential and Value-Realisation in IT-enhanced Sectors

**Gilbert Cockton**

Department of Media and Communication Design
Faculty of Arts, Design and Social Sciences
Northumbria University
UK

# WG1: Critical Reviews on Design and Evaluation Method Potential and Value-Realisation in IT-enhanced Sectors

Gilbert Cockton

Department of Media and Communication Design, Faculty of Arts, Design and Social Sciences,
Northumbria University, Newcastle upon Tyne, NE1 8ST, UK
gilbert.cockton@northumbria.ac.uk

**Abstract.** Target economic sectors are an important factor in the development of Information and Communication Technologies, but this does not apply to method transfer between sectors. Economic sectors no more relevant to method match than other factors such as key technologies, development processes or design teams' expertise with methods. The potential and effective scopes of design and evaluation methods are more complex. No single factor dominates. The scopes of development processes, methods, and the resources from which methods are formed, are better thought of broadly in terms of their coverage as *Abstract Design Situations*, which vary in the presence and balance of design choice types and their co-ordination. Options for different types of design choice do not result from methods, but from configuration and use of combined resources. It is these resources, rather than methods, that transfer between development contexts, in the same or different sectors. Resources have multiple *potential* functions that change during use across a design project, and thus resource transfer only involves potentials. The *actual* functions achieved for a resource are the result of project work, and not due to initial public forms alone. This understanding of design and evaluation work as being based on a mix of public, local and personal resources was successfully applied by researchers from outside of the action at an open research workshop in April 2013. WG1 thus created a framework for future research achievements, as well as also supporting WGs 2 and 4 during TwinTide. The main results of WG1 are listed at the end of this report.

## 1 Overview: To Be or Not To Be Methodical in Design

The nature and value of methods in Interaction Design for Information Technology have become open problems. Two challenges have proved to be insurmountable:

1) Designing robust experimental comparisons of methods that can provide reliable information to guide the selection of methods for design and evaluation work
2) Communicating methods in ways that minimise evaluator and/or designer effects

The two challenges are opposite sides of the same coin. On the one hand, evaluator and/or designer effects may make 'fair' controlled studies impossible, due to uncontrollable confounds such as participants' expertise, knowledge and specific conditions and practices during a study. On the other hand, the current manifestations of re-usable methods are often incomplete and/or ambiguous, and thus force evaluators and designers to draw extensively on their knowledge, expertise and local project resources. Both challenges thus need to be overcome together; one cannot be overcome in isolation.

The motivation to produce robust effective objective methods in Interaction Design for Information Technology arises from the initial disciplinary locations of Interaction Design and Human-Computer Interaction (HCI) in Engineering Design, i.e., Software Engineering and Engineering Psychology (also known as Human Factors or Ergonomics). In its paradigmatic ideal form, Engineering Design closely mirrors scientific method, with hypotheses replaced by requirements specifications, which represent artefacts to be implemented, as opposed to theories to be proved. Whereas experimental data in science prove a hypothesis (null or alternate), verification data in Engineering Design should establish whether an artefact satisfies its specified requirements. Similarly, although there is no single scientific method as such, but more accurately a system of values that favours some experimental procedures above others, both scientific and engineering work are expected to follow a planned method, since it is this planned method that logically allows null or alternate hypotheses to be supported (or requirements to be deliberately met and demonstrably satisfied). The design of experimental studies or rigorous engineering practice is a rational enterprise. It is the study design / work plan that allows confident acceptance of hypotheses or design success. When reporting scientific or engineering work, to allow replication or audit, the actual method must be described. If there are differences between the planned and actual method, this must not undermine the logic of the study / plan, since conclusions drawn on the basis of (verification) data may no longer be valid.

The years either side of 2000 in HCI witnessed disillusion with the realities of method use and its evaluation (e.g., Hertzum and Jacobsen 2003). Much use of supposedly well-established re-usable methods required considerable expertise, subjective judgement, and a large measure of project specific opportunism, bringing interaction design and evaluation much closer to applied arts practices in creative design. Mainstream design research had already experienced disillusion with design methods in the 1970s (Jones 1977, Cross 2001), resulting in a backlash against systematic design work in favour of designers' judgement and expertise in the context of project specific opportunities. In this sense, what was experienced in HCI research had already happened decades earlier in mainstream design research that focused on creative practices with applied arts traditions. However, the engineering heritage of HCI and Interaction Design (pre-multimedia) has resisted methodical work being abandoned as a goal, as it had been in applied arts design research. In reality, many applied arts practices never embraced design methods, resulting in unresolvable controversies between form-giving artist-designers and performance-ensuring systematic-designers that resulted in the closure of HfG Ulm, the successor of the

Bauhaus, in 1966. It was difficult for systematic methods to maintain credibility in applied arts design practices. In contrast, in engineering-oriented design practices, it has been more difficult to abandon the search for effective re-usable systematic design practices.

There are opportunities to draw on the extensive literature on design and evaluation practices in both creative design and engineering design disciplines. Rather than accept that design work is wholly a question of expertise, subjective judgement and opportunism, we can build on results from mainstream design research in HCI to provide support for design work that is not constrained by the ideals of either scientific method or the genius of the lone designer. The COST MAUSE Action prior to TwinTide made considerable progress towards such a position through the work of its Working Group 2, which focused on Comparing Usability Evaluation Methods: Strategies and Implementation. The COST TwinTide Action has built directly on this achievement.

# 2 Scoping the Work of WG1

## 2.1 Proposed Plan for WG1

The plan initially proposed for Working Group 1 (WG1), Critical Reviews on D&E Method Potential and Value-Realisation in IT-enhanced Sectors, initially comprised three tasks

1) Establish common understanding about boundary conditions for a sector by developing sector categories based on criteria such as stakeholder types, processes, products/services, technological platforms, modelling, or established tributary disciplines
2) Collect a critical mass of Design and Evaluation methods being used in a range of sectors and disciplines and create templates to describe and analyse these methods
3) Maintain a knowledge pool of analysed Design and Evaluation methods

## 2.2 Extending the Scope of WG1

At the start of the TwinTide action in Brussels in November 2009, working groups revisited the work plan for the successful proposal in brainstorming activities supported by a card sort. The divergent process here revised and extended the scope of tasks under consideration for WG1 as follows (extensions to Tasks 1-3 are indicated with italics, Tasks 4 onwards are all additions):

1) Establish common understanding about boundary conditions for a sector by developing sector categories based on criteria such as stakeholder types, processes, products/services, technological platforms, modelling, established tributary disciplines, *and communities of practice*
2) Collect a critical mass of Analysis, Design and Evaluation methods being used in a range of sectors and disciplines and create templates to describe

and analyse these methods, *making use of existing collections such as Engage (www.designandemotion.org/society/engage/)*

3) Maintain a knowledge pool of analysed Design and Evaluation methods *and case studies*, for our defined target audience
4) Document and explore alternative conceptualisations of 'sectors' and 'domains', including defining a sector and its relation/difference to: a domain, communities of practice, markets, critical aims (e.g., safety critical, high reliability), key constraints (e.g., inherent complexity), and UN SIC Codes (www.statistics.gov.uk/methods_quality/sic/contents.asp)
5) Build and prioritise a list of sectors with information that would include: relevant standards, critical issues, infrastructure dependencies and threshold concepts (www.ee.ucl.ac.uk/~mflanaga/thresholds.html), with the aim of agreeing a focus on specific sector exemplars expertise/experience within TwinTide action members, such as eLearning and social innovation.
6) Develop and maintain a knowledge pool of appropriate research methods for researching methods in use
7) Assemble a list of technologies such as Augmented Reality, noting TwinTide expertise and critical issues, that: impact on Design and/or Evaluation, underpin specific sectors (e.g., CAD in manufacturing), and their ability to transform and disrupt sectors

Given that WG1 would only be a primary focus for the first half of TwinTide, this expanded list was clearly ambitious, but it gave action members stronger initial ownership of the WG1 tasks. This extended task list was consolidated during discussions at the First MC meeting in Newcastle in March 2010. The main decision here was to focus on the last three tasks above (5-7: all added at Brussels) to better support the first four. The overall preliminary aim was thus to

> *Consider the influence of sectors, technologies, research methods and human roles in user experience work in Interaction Design.*

Despite TwinTide's focus on sectors as a key influence on method transfer, WG1 members argued for the need to consider other factors, and to not presume that sectors were always the primary influence. Technologies, research expertise, preferred development processes, and method innovation (especially '3rd Wave HCI methods' with a user experience focus) were no less relevant in any initial analysis.

### 2.3 Sectors and other important factors in user experience work

Most formal scheduled WG1 activities occurred between October 2010 (Reykjavik meeting) and April 2011 (Limassol meeting). From late 2011 onwards, WG1 work progressed through STSMs and workshops at conferences and training schools.

Before the Reykjavik meeting, we surveyed TwinTide action members' familiarity and expertise with the following factors:

1) Sectors (lead Gilbert Cockton)
2) Technologies (lead Nikos Avouris, University of Patras)

3) Research Methods (lead Jettie Hoornhout, Philips Research)

The leads for each factor collated members' responses and an open card sort was prepared. In an open card sort, cards with individual factors on (e.g., agricultural sector, virtual reality technologies, data analysis research methods) are grouped by similarity, with further high level groups also formed and names until top level categories are reached. Cards sorts for each factor were carried out by three separate groups. Each group reported back on their results, which were too disjoint to allow single taxonomic structures for each factor. However, discussions on differences of approach were very productive.

On *Sectors*, each group's card sorts produced structures similar to the top and first level categories of the United Nations' International Standard Industry Classification (ISIC, United Nations 2008). This is an economics resource for national statistics reporting, with a focus on inputs, processes and outputs, which can be grouped by high level sectors from Economic Geography (e.g., Primary, Secondary, Tertiary etc.). However, it was observed that sectors can also be understood culturally as having distinct values (e.g., a marketing analysis could identify shared brand values for a sector, as could business strategy analysis). Whatever the relationship between sector values and modes of production, it is clear that sectors differ in terms of what each valued .

On *technologies*, a common organising structure was based on abstract system architectures, with technologies grouped by input, processing, storage, communication and output. By distinguishing between different stages of processing, this results in a structure similar to Garrett's (2010) Elements of User Experience, which in turn is close to the abstract architectures for interactive systems of the 1990s.

On *research methods*, many members had extended their lists of expertise to include a wide range of design and evaluation methods, and not only those such as questionnaires or interviews that would commonly be covered in a human sciences text book on research methods. For this factor, a common organising structure was based on phases or stages of software development processes.

Together, these analyses and discussions extended the scope beyond TwinTide's initial focus on sectors to a broader consideration of factors that could have a bearing on method transfer. However, the Limassol meeting (April 2011) refocused on sectors to explore the relationship between sectors and values. The motivating conjecture here was that the methods are also guided by values, and thus the transferability of a method would be dependent on the match between method and sector values (Cockton 2011), anticipating the exploration of diffusion of innovation perspectives within TwinTide WG4 from 2012 onwards. Match between benefits and needs is a key factor in diffusion of innovation.

At the Limassol meeting, action members split into groups and chose a sector where one or more members had specific expertise (as earlier indicated for the Reykjavik surveys). Each group brainstormed on core values for their chosen sectors. They also brainstormed on key technologies for a sector, to continue the broader focus from Reykjavik. The main outcome of this exercise is that sectors were not sufficiently distinct to result in radically different matching criteria for method diffusion through intersector transfer. While not all sectors were associated with

identical values, there was extensive overlap. This was consistent with much work on business strategy, where many key factors are cross sector. For example, in the Balanced Scorecard approach (Norton and Kaplan 1996), targets and objectives are set from the customer's and three company perspectives: financial, learning and growth, and internal business processes. Competitive strategies should set targets and objectives that result in a balance of benefits between stakeholders. This approach thus considers value from the perspective of the consumer and the company. Financial worth is largely a function of return on investment. Internal business processes can be made more efficient and effective. Learning and growth within organisations relates to a third set of strategic values that are largely cross sector, as is the fourth factor of customer satisfaction. These four strategic perspectives (financial, customer, internal business processes, learning and growth) all connect with cross sector values.

The insight that there was extensive overlap across sectors' values originated with a group considering agriculture, where labour force development in an east European country is a primary value driving contemporary development. This was clearly a value relevant to all sectors, indicating, for example, that all economic sectors can benefit from e-learning technologies for workforce development, and thus any design and evaluation methods for adult e-learning may meet a sector's needs. All organisations need to value learning and growth, regardless of economic sector.

Once again, sectors were not seen as an inherently dominant factor in method transfer, and were not necessarily more relevant to method match than other factors such as key technologies, development processes or design team method expertise. The interactions between methods and application contexts were more complex. The scope of methods was not simply one of sector match, but a broader one involving not only sector values, but also system technologies, user groups and development processes (especially evaluation preferences and budgets). Such a scope closely matches Abstract Design Situations (Cockton 2010), a formalisation of Heskett's (2002) observation that design outcomes result from four types of choices, which are ones about means (e.g., technologies), about ends (e.g., sector values), about beneficiaries (e.g., user groups), and about how design choices are evaluated.

An Abstract Design Situation reduces complex concrete design settings to co-ordinations of design choice types. Abstract Design Situations vary in the presence and balance of design choice types and their co-ordination. Clearly, all design settings involve choices about the artefacts being developed (Heskett's 'means'), but they vary in the effort allocated to clearly stating design purpose (Heskett's 'ends'), to developing understandings of beneficiaries, and to evaluations of the extent to which artefacts meet the purpose for which they were designed, when used by intended beneficiaries.

The scopes of development processes, methods, and the resources from which methods are formed, are thus better thought of broadly in terms of their coverage as *Abstract Design Situations* (ADSs), and not narrowly in terms of sectors and their values that guide design purpose. This understanding was carried forward to WG1 activities in the remainder of TwinTide.

Development processes were not considered as scoping methods at Limassol. While some groups at Reykjavik used development stages to group methods, it is possible to use many methods and their outputs at several points in a project, especially ones that do not follow engineering project management models and do not have homogeneous stages that focus exclusively on problem analysis, requirements specification, conceptual design, detailed design, implementation, verification and validation, installation and operation.

The concept of Abstract Design Situations was developed by WG1's leader (Cockton 2010) before the Reykjavik workshop, but it was not recognised as a better basis for scoping design and evaluation methods, in comparison to sectors alone, until 2013 during the consolidation of insights from STSMs and conference workshops where WG1 work had been shared with a range of international audiences (Cockton 2011, Cockton 2012, Cockton 2013a, Cockton 2103b, Cockton 2013c). This is covered in more detail in a later section of this report.

## 2.4 Consolidation

After April 2011, the focus of TwinTide activities shifted to WGs 3 and 4. An exception here was the CHI 2013 workshop in Paris (Law et al. 2013), which jointly focused on WG1 and WG4 activities. While WG1 activities from 2009 to 2011 involved all members of the TwinTide action, WG1 work in the second half of the action was carried out by a small group of interested members and their PhD students, and thus most references to this work are (co-)authored by the WG1 Leader and author of this report.

In September 2011, WG1 and WG2 activities were aligned at a WG2 workshop in Linz. WG2 had explored conceptualisations of method and sector in Reykjavik and Limassol using knowledge elicitation methods. This had resulted in rich characterisations of both, in contrast to WG1's acceptance of the UN (2008) ISIC framework as the basis for understanding sectors. As for methods, WG1 had continued MAUSE WG2's understanding of methods as achievements, i.e., the actual methods that are constructed in use, rather than methods as premonitions, i.e., public methods that provide complete plans before the start of creative design or related evaluation work. The scientific ideal of fully planning all procedures in advance is difficult to achieve in design work, where expertise and opportunities shape evolving practices. Instead, what pre-exists design work are approaches that are loose collections of resources with different configuration needs.

A design project can begin with a set of approaches in mind that will be used at different points in development. However, the resources associated with an approach are incomplete, and have to be completed (e.g., participant recruitment plans for user testing), adapted (e.g., existing persona skeletons) or complemented (e.g. user testing methods do not provide data gathering methods, these need to be chosen and added to user testing plans). MAUSE WG2 argued for understanding design work at both project and resource levels, rather than at method level, which has dominated design research since the 1960s (Cross 2001). Design research thus needed to both broaden

and narrow in scope: broadening out to consideration of design work holistically, and narrowing down to understanding the dynamics of resource use within design work.

MAUSE WG2 had provided an initial typology for evaluation resources (Woolrych et al. 2011):

1) Scoping (understood by the Linz meeting as corresponding to an Abstract Design Situation)
2) Axiological (values motivating a resource)
3) Instrumentation (for data collection)
4) Procedural (the directive skeleton of a method)
5) Expressive (the intermediate and final results of a method)
6) Knowledge (the expertise underpinning a method)
7) Process (scope within a development process)

During WG2 work at Linz, the concepts in use for case study analyses (initiated at Limassol) were mapped against these resource types, and participants' brainstorming was observed. As a result of this, the MAUSE resource types were revised. *Instrumentation* resources were generalised beyond evaluation to become *harvesting* resources. Procedural and process resources were merged to become *directive* resources, resulting in six resource types, which were now aligned with the decision after Limassol to not see development process phases or stages as a reliable scoping structure.

Resource type analysis was used, along with Abstract Design Situations, with PhD students at the November 2011 TUTOREM PhD training school at Bertinoro in Italy (this repeated an exercise from the German DGTF 2011 conference, Cockton 2013c). The results were positive, and were reported to action members in the MC meeting after the PhD School, along with the outcomes of the Linz workshop that were relevant to WG1. As there would be no scheduled WG1 activities during 2012 management meetings, the focus returned to the first three tasks agreed at the Brussels meeting, which were simplified as follows:

Task 1: Define 'Sector' in support of other WGs (merging first and fourth Brussels tasks)
Task 2: Characterise the nature of methods
Task 3: Identify Design and Evaluation methods in terms of approaches and resources, scoped relative to Abstract Design Situations

Task 1 was completed by Bertinoro. Sectors of interest were as analysed at Limassol, and these remained in scope for WG3 and WG4 tasks for the remainder of the TwinTide project. The completion of Task 1 drew on the work of the UN (2008) Department of Economic and Social Affairs Statistics Division as follows:

- economic production is an activity, carried out under the responsibility, control and management of an institutional unit, that uses inputs of labour, capital, and goods and services to produce outputs of goods and services
- a production-oriented or supply-based conceptual framework groups producing units into detailed industries based on similarities in the economic activity, taking into account the inputs, the process and technology of

production, the characteristics of the outputs and the use to which outputs are applied. The weights assigned to these types of criteria may vary from one category to another and between different levels of the classification.

WG1 therefore defined a sector as:

- a producing unit with common inputs, process and technology of production, common characteristics of outputs and common uses to which outputs are applied.

Values, which had become a focus of initial WG1 work, are associated primarily with the common uses to which outputs are applied, but characteristics of outputs and common inputs, process and technology of production also matter when considering sector requirements. However, except where institutions provide services to individuals, families, and/or communities, this definition is economic and thus excludes interaction and software design for:

- Individuals (self, identity, growth, health and wellbeing ...)
- Families (co-ordination, communication, caring, ...)
- Communities (information, sharing, support, issues, ...)

Task 1 (Define 'Sector' in support of other WGs) was thus completed at Bertinoro. However, by that point in the TwinTide COST Action, this was of limited value, since sectors had been seen to be an inappropriate basis for establishing the scope of design and evaluation methods. Furthermore, the economic focus of a Task 1 defined based on UN ISIC would only allow design and evaluation methods to be scoped relative to design purpose for institutions, and not for other stakeholder types (e.g., individuals, families, communities). An economic sector focused approach to design work thus inevitably excludes some key application areas for interaction design.

Fortunately, Task 3 *Identifying Design and Evaluation methods in terms of approaches and resources*, scoped methods in terms of Abstract Design Situations (ADSs), which became the main focus in the closing stages of TwinTide WG1, since this was a more general scoping construct than economic sectors.

Informal activities after the Limassol WG1 activities advanced Task 2 (Characterise the nature of methods). The position consolidated at Bertinoro was as follows:

- Methods are post-hoc rationalisations of design work
  - Practice precedes abstraction
  - Method = Abstraction of Practice
- Practice combines:
  - Re-usable ('pre-figured') public resources that are generic, or are part of named approaches. An example of a generic resource is published usability problem report forms.

- o Local project , organisational and team resources
- o Personal individuals' resources

- Pre-figured resources may be combined into named approaches.
- Practice = the use of public, local and personal resources through selection, provision, adaptation and innovation

Examples of approach resources for the named approach of Heuristic Evaluation are:

- o Heuristics (knowledge resources)
- o inspection procedures (directive resources)

and for user testing:

- o test procedures (directive resources)
- o data collection instruments (harvesting resources)

Examples of local project resources include:

- o Participant selection criteria and screening forms
- o Local problem report and informed consent forms
- o UX experts' knowledge

It is thus resources, rather than methods, that transfer between development contexts, in the same or different sectors. Such resources may be grouped into named approaches, but the latter will not provide completely prefigured methods that can be used without extra configuration work. Instead, specific public approach resources may be incomplete, often to a great extent, and even where an approach's public resources are each individually complete, local resources may still be needed to complete a method based on an approach's public resources.

# 3 Remaining Work on WG1

There was no WG1 work at formal TwinTide management meetings in 2012, as all time was given over to WG3 and WG4 work. However, four STSMs for PhD students at Northumbria (UK) and Delft (NL) were used to test the resource type framework as revised at Linz and reported at Bertinoro. This work delivered significant insights that corrected an oversimplification from the MAUSE COST action. Instead of design and evaluation resources being typed, instead

> *Resources have multiple potential functions that change during use across a design project. Resources that have multiple dynamic functions cannot be given single types.*

This work extended consideration of resources to ones that are not part of approaches beyond generic resources such as evaluation report formats. A range of local resources are independently introduced in design settings and may not become part of a completed method, but instead have process level functions that contribute to social and emotional aspects of design work.

Preliminary results from STSMs were reported at a WG4 workshop in Delft (September 2012) with updates at the MC meetings in Copenhagen (October 2012) and Coimbra (March 2013). A final overview was presented at the MC meeting in Tallinn (July 2013). Full results will be presented in the PhD theses and publications of Northumbria and Delft students supported by TwinTide STSMs.

These results informed a successful proposal for a CHI 2013 workshop, *Made for Sharing: HCI Stories of Transfer, Triumph & Tragedy*, held in Paris in April 2013 (Law et al. 2013), that confirmed the viability of an extended set of resource functions arising from the STSM work. Workshop paper authors were able to respond to the understandings of method resources as communicated in the workshop's call for participation, in ways that are clearly indicated in some accepted paper titles and abstracts at http://www2.le.ac.uk/departments/computer-science/people/elaw/HCI-3T. The understanding of design and evaluation work as being based on a mix of public, local and personal resources developed within the TwinTide action thus could be applied by researchers from outside of the action.

The extended set of resources was presented via a range of alternative vocabularies in an experimental alt.chi paper at CHI 2013 in Paris (Cockton 2013b). Three are shown in Table 1 below. Referring to the right column of the table, expressive functions became limited to private use by design team members. Resources for communication beyond professional designers or UX specialists have a newly identified *performative* function.

Three further new resource functions were identified during applications of TwinTide's theorization of design work. *Invigorative* and *protective* functions support emotional aspects of design work. *Integrative* functions co-ordinate across design activities within a development process. These four additional functions extended the MAUSE COST project's understanding of resource use from evaluation work.

| **Everyday Vocabulary** | **Technical Vocabulary** | **Challenging Vocabulary** *(with helpful glosses)* |
|---|---|---|
| Limiting | Utilisation | Adumbrative *(rough outline of scope)* |
| Valuing | Prioritisation | Ameliorative *(guiding values)* |
| Sourcing | Investigation | Inquisitive *(finds stuff out)* |
| Steering | Instruction | Directive *(systematically guides design)* |
| Recording | Registration | Expressive *(gets stuff down)* |
| Telling | Education | Informative *(puts stuff in)* |
| Sharing | Presentation | Performative *(spreads stuff out)* |
| Energising | Acceleration | Invigorative *(spurs things on)* |
| Caring | Correction | Protective *(keeps things up)* |
| Linking | Co-ordination | Integrative *(pulls stuff together)* |

**Table 1**. Alternative vocabularies for a set of resources

Resource functions have become the dominant construct for understanding re-usable aspects of software design and evaluation work. The initial *Working to Choose* (W2C) framework developed for the DGTF 2011 workshop (Cockton 2013c) was

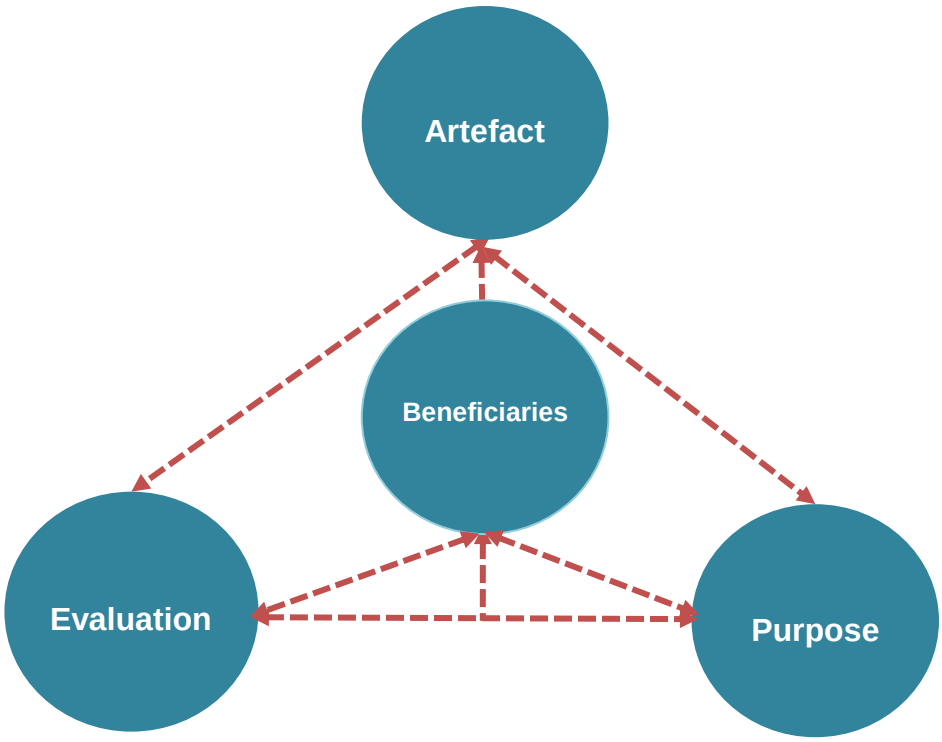



**Fig 1**. The structure of an Abstract Design Situation

refined for the CHI 2012 workshop (Cockton 2012), but it gave equal status to resource functions, Abstract Design Situations (ADSs) and a further concept of Meta-Principles for Designing (Cockton 2009). The third concept had been reduced to an evaluative role in (Cockton 2012). ADSs are now the structuring concept for scoping functions (limiting, utilisation, adumbrative in table above). They are thus subordinate to resource functions, greatly simplifying the W2C framework for analysis of design work.

The use of ADSs to scope approaches and resources gives us a broader basis for understanding method transfer, the key focus for the TwinTide COST action. Heskett (2002) noted that design outcomes result from choices of means (artefacts), ends (purpose), beneficiaries (users) and evaluations.

An ADS (Cockton 2010) is characterised by the balance and coordination of up to four different types of design choice: artefacts, purpose, beneficiaries and evaluations. Methods are thus similarly scoped by technologies (artefacts), values (purpose), sector roles (beneficiaries) and evaluation requirements. Again, method match is not wholly a question of sector values. Instead, sector differences may be reflected in all four design choice types.

Forming a method for any new ADS is thus a form of transfer (as new to world or to a design team). The novelty of a method reflects the distinctiveness of the ADS that scopes it, i.e., in terms of its applicability (transferability) to new technologies, new features, new benefits, new costs, new risks, new age groups, new impairment groups, new cultures, new quality criteria, new measures, new evaluation instruments etc. Figure 1 shows the structure of an ADS (choice type names above, matching scopes below on each circle).

This simplification of W2C in the context of the position reached at Bertinoro in 2011 marked the completion of Task 2 (*Characterise the nature of methods*). Task 3 (Identify Design and Evaluation methods in terms of approaches and resources) is being continued in ongoing PhD work. This could not be completed within the time frame of TwinTide as the analytical framework developed within Task 2 was not completed in time. This PhD work at Northumbria (Jones et al. 2012a, Jones et al. 2012b, Leitner et al. 2013) builds on group discussions at the CHI 2013 workshop, which identified some key issues associated with resource based analysis of design work. One such key issue is:

> *How do resources and their dynamic functions manifest themselves in design work?*

The holistic nature of design work makes it difficult to track the introduction, evolution and use of design and evaluation resources. One of the PhD students in the Delft and Newcastle STSMs, Malcolm Jones, has developed innovative storyboarding approaches to tracking resource life cycles within design work.

A future second key issue involves how to transfer the insights from academic experts on resource function analysis to practising designers and researchers who apply, rather than create and/or analyse HCI design resources. This could be achieved via changes in the ways that re-usable approaches and resources are published and disseminated.

A third key issue involves the role of resources that are not part of approaches. Ongoing PhD work has shown that resources can be process level ones with social, emotional or integrative functions, and not necessarily only within named approaches. This has expanded the MAUSE scope beyond cognition. Also, design processes are not just formed from completed methods, but instead also involve generic resources that exist independently of methods. These are often social (performative), emotional (invigorative, protective) or integrative. They are often spontaneously triggered. Transfer and diffusion may require new or adapted resources at both approach and process level.

Thus, while Task 3 did not complete as planned within the time frame of the TwinTide COST action, Tasks 1 and 2 created a framework for future research achievements, and also supported Working Groups 2 and 4 within TwinTide. WG2 benefitted from a common understanding of the nature of methods and transfer. WG4 was able to develop understandings of design and evaluation methods as local completions of approaches' resources thus led to the Diffusion of Innovation focus, since it is through local completion that methods are diffused, and the matching process includes both assessment of approaches as published and assessment of the local costs of configuration and completion.

# 4 Achievements of WG1

WG1 has achieved the following main results:

- Application of value and technology analyses to UN ISIC sectors in IT development, forming a wider basis for scoping approaches and resources.
- Formulation of broad complex scopes for approaches and resources, which can be mapped onto Abstract Design Situations (Cockton 2010) from design theory, but reflects the existing expertise of TwinTide action members in refining and evolving the initial task structure for WG1 beyond the proposal's focus on sectors
- Substantial refinement of the resource types from MAUSE WG2, applying them to TwinTide WG2 work, extending them from usability methods to design and evaluation methods in general, and reconceptualising types as *functions* that can be *multiple* and *dynamic*, *evolving* and *changing* as design processes progress
- Development and initial validation of an analytical framework based on resource function analysis to the manifestation and use of resources in live design workshops (achieved through STSMs and CHI 2013 Paris workshop).
- Refinement of a framework for analysing, evaluating, improving and communicating re-usable resources and approaches for design and evaluation work, including the diffusion of method innovation (WG4 work)

The WG1 framework is now in use with PhD students and researchers in the UK, Finland and Poland, and was applied in part by participants at the CHI 2013 Paris workshop. It offers a middle ground between the scientific values of idealised engineering design and the chaotic romanticism of extreme creative design practices. It provides a basis for studying the impact of re-usable resources and approaches in realistic design settings. It sets realistic expectations for re-usable design support and sets new standards for how this is communicated to potential beneficiaries. It has already guided development of innovative design resources within TwinTide STSMs that are already being recognised as internationally outstanding (e.g., Michael Leitner's honorable mention, one of five only, for his Mobile HCI 2013 paper Leitner et al. 2013).

The conceptual framework associated with resource function analysis continues to attract interest across Europe from researchers and practitioners. It has been widely disseminated through publications, conference workshops and keynote addresses from 2011 onwards. Thus while some of the initial aims of WG1 were not achieved, a novel framework with demonstable value resulted from its work.

## 5. References


1. Cockton, G. (2009) Getting There: Six Meta-Principles and Interaction Design, in Proc. CHI 2009, eds. D.R. Olsen Jr., R.B. Arthur, K. Hinckley, M.R. Morris, S. Hudson & S. Greenberg, ACM, 2223-2232.
2. Cockton, G. (2010) Design Situations and Methodological Innovation in Interaction Design" in CHI 2010 Extended Abstracts (alt.chi paper), eds. E.D. Mynatt, D. Schoner, G. Fitzpatrick, S.E. Hudson, K. Edwards and T. Rodden, 2745-2754
3. Cockton, G. (2011) Domain Values and Method Transferability: an Initial Framework, 1st European Workshop on HCI Design and Evaluation, eds. G Christou, P. Zaphiris and E.L-C. Law, 85-90, IRIT Press, Toulouse, France, ISBN :978-2-917490-13-6
4. Cockton, G. (2012) Making Designing Worth Worth Designing. Position Paper, CHI 2012 Workshop on Methods for Accounting for Values in Human-Centered Computing. Available from http://ii.tudelft.nl/ValuesInDesign/submissions/cockton.pdf
5. Cockton, G. (2013a) Diffusion of Worth Mapping: The Worth of Resource Functions, position paper for CHI 2013 workshop, Made for Sharing: HCI Stories of Transfer, Triumph & Tragedy. Available from http://www2.le.ac.uk/departments/computer-science/people/elaw/HCI-3T/P2Cockton.pdf
6. Cockton, G. (2013b) A Load of Cobbler's Children: Beyond the Model Designing Processor, alt.chi paper, In CHI '13 Extended Abstracts on Human Factors in Computing Systems (CHI EA '13). ACM, 2139-2148.
7. Cockton, G. (2013c) You (have to) Design Design, Co-Design Included, in Mareis, C., Held, M., and Joost, G. (eds): Wer gestaltet die Gestaltung? Praxis, Theorie und Geschichte des partizipatorischen Designs. Bielefeld: transcript. 181-205, ISBN 978-3-8376-2038-2, invited paper presenting the content and outcomes of workshop for junior researchers (invited workshop leader)
8. Cross, N. (2001). Designerly ways of knowing: design discipline versus design science. Design Issues, 17(3), 49 55.
9. Garrett, J. (2010).The Elements of User Experience: User-Centered Design for the Web. New Riders publishing
10. Hertzum, M. & Jacobsen, N.E. (2001). The Evaluator Effect: A Chilling Fact About Usability Evaluation Methods. IJHCI, 13(1), 421-443.
11. Heskett, J. (2002). Toothpicks and Logos: Design in Everyday Life. Oxford University Press
12. Jones, J C. (1977) How My Thoughts About Design Methods Have Changed During the Years, Design Methods and Theories, 11(1).
13. Jones, M, Teinaki, V.and Leitner, M. (2012a) Supportive Resources: Mastering the Art of Making Sense. In: Message in a Bottle DIS 2012 Workshop, 12 June 2012, Newcastle University.
14. Jones, M, Teinaki, V.and Leitner, M. (2012b) Supportive Resources, Interfaces Magazine 91, 15-17
15. Law E. L.-C., Hvannberg E., Vermeeren A.P.O.S., Cockton G., and Jokela T. (2013) Made for Sharing: HCI Stories of Transfer, Triumph & Tragedy. In CHI

'13 Extended Abstracts on Human Factors in Computing Systems (CHI EA '13). ACM, 3235-3238. All workshop papers are available via http://www2.le.ac.uk/departments/computer-science/people/elaw/HCI-3T/accepted-submissions

16. Leitner, M., Cockton, G. and Yee, J.S.R. (2013) At the mobile experience flicks: making short films to make sense for mobile interaction design. In Proceedings of the 15th international conference on Human-computer interaction with mobile devices and services (MobileHCI '13). ACM, 304-307. Honourable Mention Award.
17. Norton, D.P. and Kaplan, R.S., (1996) The Balanced Scorecard: Translating Strategy into Action, Harvard Business School Press.
18. United Nations, (2008). International Standard Industrial Classification of All Economic Activities, Department of Economic and Social Affairs Statistics Division, Revision 4ST/ESA/STAT/SER.M/4/Rev.4, UNITED NATIONS PUBLICATION, Sales No. E.08.XVII.25, ISBN: 978-92-1-161518-0
19. Woolrych, A. Hornbæk, K. Frøkjær, E. and Cockton, G. (2011). Ingredients and Meals Rather Than Recipes: a Proposal for Research that Does Not Treat Usability Evaluation Methods as Indivisible Wholes, in IJHCI, 27(10), 940-970


# TwinTide WG2 Report

## Transferability of Design & Evaluation Methods across IT-enhanced Sectors

**Christian Stary & Matthias Neubauer**

Johannes Kepler University Linz
Austria

# WG2: Transferability of Design & Evaluation Methods across IT-enhanced Sectors

Christian Stary & Matthias Neubauer

Johannes Kepler University Linz, Department of Business Information Systems - Communications Engineering, Science Park 3 Altenbergerstraße 69, 4040 Linz, Austria, christian.stary@jku.at; matthias.neubauer@jku.at

**Abstract.** Working group 2 (WG2) of the IC0904 TwinTide action has focused on the transferability of Design and Evaluation Methods across IT-enhanced Sectors. This final report describes the working group objective and the methodological approach used by WG2 to reach this objective. A method-mix using methods from social-sciences and knowledge management has been applied in order elicitate, represent and negotiate expert knowledge on method transfer. In this report the method-mix and selected results are presented. Furthermore, impacts achieved in terms of publications and outreach activities are listed and a general conclusion on method transfer and future work is given.

## 1 Objectives of WG2

The primary objective of working group 2 was to stimulate and integrate research on design and evaluation methodologies. A more specific focus was the identification and analysis of sector-unique design and evaluation methodologies in terms of their underlying conceptual models, practical protocols, and transferability to other sectors.

To achieve this objective, three major tasks were dealt with:

1. Identify criteria for assessing method transferability;
2. Collect cases of successful and non-successful attempts to transfer design & evaluation methods across sectors and analyze the associated contributing factors;
3. Analyze design and evaluation methodologies commonly applied uniquely in one sector and across a range of sectors.

Within the following sections the methodology of WG2 to perform the tasks as well as the results of WG2 will be presented.

## 2 Approaching Transferability of methods

WG2 made use of user and task knowledge as focal points for transferring Design & Evaluation methods. Starting with a Concept Analysis, the members of WG2 aimed to develop a common understanding of the construct ‘method’. Concept Analysis is

an established technique used in the social sciences with the purpose of deriving a ‘formula' that in turn can be used to generate definitions and descriptive phrases (cf. Dalkir, 2005). This approach is particularly useful in tackling multidisciplinary domains such as HCI, in order to develop clear criteria to enable sorting into categories such as design versus evaluation, tasks versus goals, or usability versus user experience. In addition, valuable contributions to the project can be derived through production of ontologies (semantic maps of key concepts), identification of core features of methods, and identification of know-how, and know-why.

Concept analysis can be used to visually map out conceptual information in the process of defining a term. This technique can be used in clearly defining complex, value-laden terms such as refinement, tuning or abstraction. It is an approach to help develop a rich, in-depth understanding of a concept. Concept analysis rests on obtaining consensus on three major dimensions of a given concept (Table 1):

1. A list of key attributes that must be present in the definition, vision, or mission statement.
2. A list of illustrative examples.
3. A list of illustrative non-examples.

| **CONCEPT NAME:** | **‘METHOD’** | |
|---|---|---|
| *Attribute* | *Example* | *Non-Example* |
| Rationale | Find usability problems | Monitoring of work conditions |
| Procedure | Steps to take to create a UML diagram | UML diagrams themselves |
| etc. | etc. | etc. |

**Table 1.** Concept Analysis for the construct ‘method‘

Method attributes were individually collected within a Brainstorming session (cf. Rawlinson, 1981) and written on paper cards. Following, the collected attributes have been consolidated and grouped by the dimensions ‘Why’, ‘What’, ‘How’ and ‘What Else’ by the WG members. The results of this first WG session were documented within the TwinTide Wiki which enabled collecting examples and non-examples for each attribute between the WG meetings (cf. Table 1).

The results of the first WG meeting and the online activities considering the concept analysis were consolidated within the second WG meeting in a group discussion. Detailed results are provided in the Appendices section ‘Concept Analysis for the contruct method’ Based on a common understanding of the construct ‘method’ and its attributes, WG2 members continued with the discussion of the transferability of methods. In advance to the discussion a literature review was conducted by the WG leader and results (cf. Bowman, Gabbard and Hix, 2002; Fitzpatrick, 1999; Introna and Whitley, 1997; Steves, Morse, Gutwin, Greenberg, 2001; Szulanski, 1996) were

presented to the members of WG2. The results indicated three different, relevant aspects considering the transferability of methods

- Application context
- Methodological context
- Transfer context

These aspects served as basis for discussing method transfer within a question café setting which is similar to a world café (cf. http://www.theworldcafe.com/):

- Café tables are set up (provide paper + markers)
- To each table a concrete question is assigned. In the setting of WG2 the following questions were assigned:

  (a) What do we have to ask to understand project specific factors (application context)?
  (b) What do we have to ask to understand method specific factors (method context)?
  (c) What do we have to ask to understand transfer specific factors (transfer context)?

- at each table a responsible host is defined
- the members discuss at a table about 20min and afterwards switch to another table
- after a number of "cafés" the results at each table are presented by the hosts

Core findings of the question café are listed below:

- How much experience/knowledge is there of using the method successfully in current domain?
- What are domain specific factors? In what respect do the domains differ (e.g. type of products, development process, target users, environment,…)
- Are there related methods in the new domain?
- What added value would the transferred method provide in the new domain?
- How much time and expertise is available for the transfer?
- Once transferred, can the method be used? (expertise, resources)
- What key attributes can be changed?
- What key-attributes are fixed?
- What key attributed should be changed?
- When is it an adapted method vs. a new method?
- Are there cultural factors that determine particular methods?
- What 'values' are embodied in or communicated through a method?
- Are there laws or restrictions for being allowed or having to use specific methods?
- Is the application safety or business critical?
- Which development processes does the method fit with?
- How reliable is the method?

A complete list of questions derived from this session which is relevant for the discussion of Transfer Case Studies is provided within the Appendices (cf. 'Towards Transferability – Results of the Question Café in Reykjavik'). Besides the questions, the rationale for each question is given. The actions described so far allowed to gain knowledge on method attributes and transferability and they contributed to perform task 1: 'Identify criteria for assessing method transferability'. Based on this knowledge WG2 proceeded with the validation of the knowledge within concrete method transfer cases. Doing so, the members of WG2 investigated case studies (e.g. Blandford, 2006, 2007) considering method-specific, application-specific and transfer-specific (meta-)information (cf. Fig. 1). The results from the case studies confirmed the knowledge gained beforehand. However, transfer-specific information was hardly provided within the cases. Mainly method- and application-specific (meta-) information was retrieved from the cases, e.g.:

- *Method-specific (meta-)information:* Formality, Complexity, Learning curve, Time to perform, Objectives, Tools to apply, Procedure.
- *Application-specific (meta-)information*: Applicability, Willingness to share knowledge about the application of a method within a given sector, Available expertise, Domain knowledge, Users.

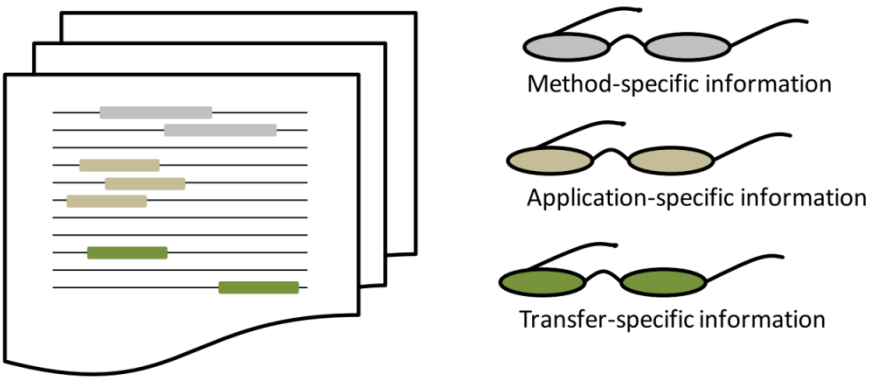



**Fig. 1.** Case Study Analysis

Within a result alignment meeting WG2 members reviewed existing results and consolidated the key attributes in terms of their relevance within transfer situations and their meaning. In an interactive workshop the key attributes have been sorted with respect to their relevance in transfer situations. Subsequently, relationships between attributes have been identified. The results of this workshop are depicted in figure 2. Green card are used for relationships, yellow cards represent method attributes relevant in the context of transfer and white cards depict method attributes that have not been considered relevant when transferring a method.


**Fig. 2.** Consolidation of Method attributes and their relationships in the context of transferability

In order to support online discussions and knowledge exchange, the results of WG2 have been documented within the Nymphaea platform (User/Password: twintide available at: https://nymphaea.ce.jku.at/nymphaea/). Nymphaea can be considered as knowledge management system, as it supports individual and collective documentation and exploitation of knowledge. Hereby, content is structured according to metadata. These can stem from domain ontologies or be specified for a special purpose (as in our case). This metadata based content management serves as baseline for cognitive and social interaction. So-called views allow annotating the tagged content with comments, remarks and links. The latter either refer to other content elements or to social media entries, such as chats, blogs, or entries in dedicated community media.

Results from the consolidation (cf. Fig. 2) have been documented as concept map (cf. Fig. 3) in the Nymphaea platform. The concept map represents a "transfer template" that supports describing the application of methods as well as the transfer of a method. The template incorporates transfer-specific information items that have been considered relevant when reconsidering the key attributes gained from the concept analysis. It can serve as basis for structuring

- Case studies for the application of a method in different domains
- Case studies for the transfer of a method to another sector (application domain)

- Case studies for the development of a new method for an application domain

Using the template a common navigation structure gained from expert knowledge can be provided. It aims to support practitioners as well as experts in the course of method transfer, application and development. Following the template is depicted as concept map in Fig. 3. The map includes key attributes of a method considered relevant in the course of change (blue and yellow concepts; blue concepts are "meta-concepts" that can include yellow concepts, red concepts have a low priority when talking about transfer). The identified and shared clusters within the map are:

- Origin, such as theoretical underpinnings
- Procedure – some need to adapt to context others
- Resources, such as required experts
- Body of knowledge, such as examples of applications

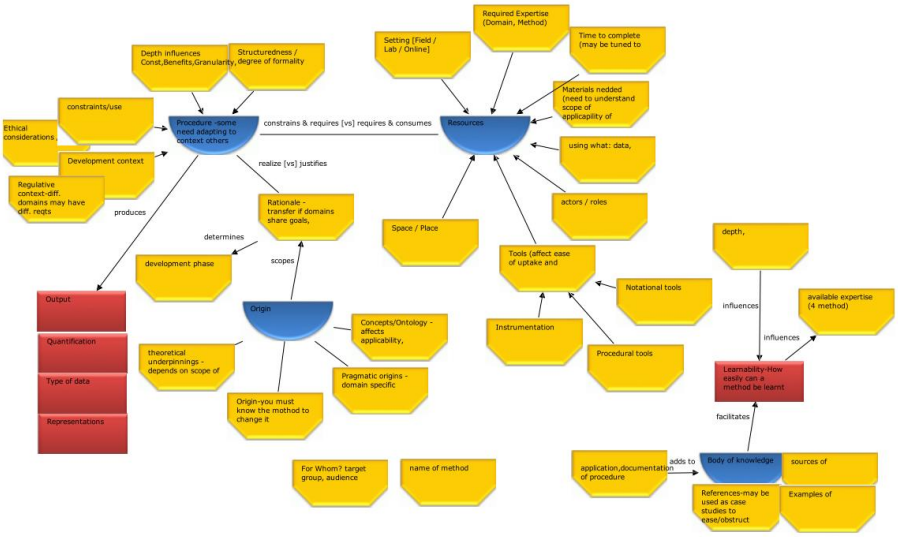



**Fig. 3.** Transferability Map

Utilizing the Nymphaea system case studies can be structured and enriched with metadata (cf. "directive" within figure 5) relevant in the context of method application and method transfer. Users have the possibility to annotate information (cf. Fig. 4) they consider application-specific, sector-specific or transfer-specific with different colors. Beyond, support for sharing and discussing their insights with other participants is provided due to public annotations and case specific forums (cf. Fig. 7).

Fig. 4 depicts the (hierarchical) structure of the case study "Claims analysis in the wild" (Blandford, 2006) and the selected content for applying claims analysis. Using metadata such as 'Directive' users can filter content accordingly (cf. Fig. 5). Besides,

provided content can be individually annotated (cf. highlight, link, note) and moreover discussed in a context-sensitive way by directly linking content elements to fora entries (cf. Fig. 7). Fig. 6 shows the alternative navigation using a part of the transfer template, where concepts are linked to content items and again the links can be enriched with transfer relevant information.

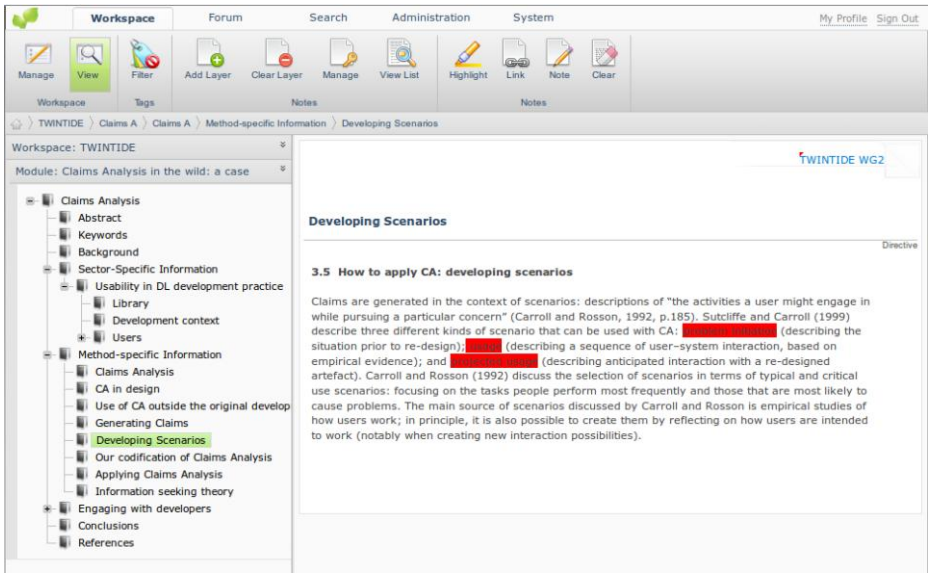



**Fig. 4.** Enrichment of content in the Nymphaea platform

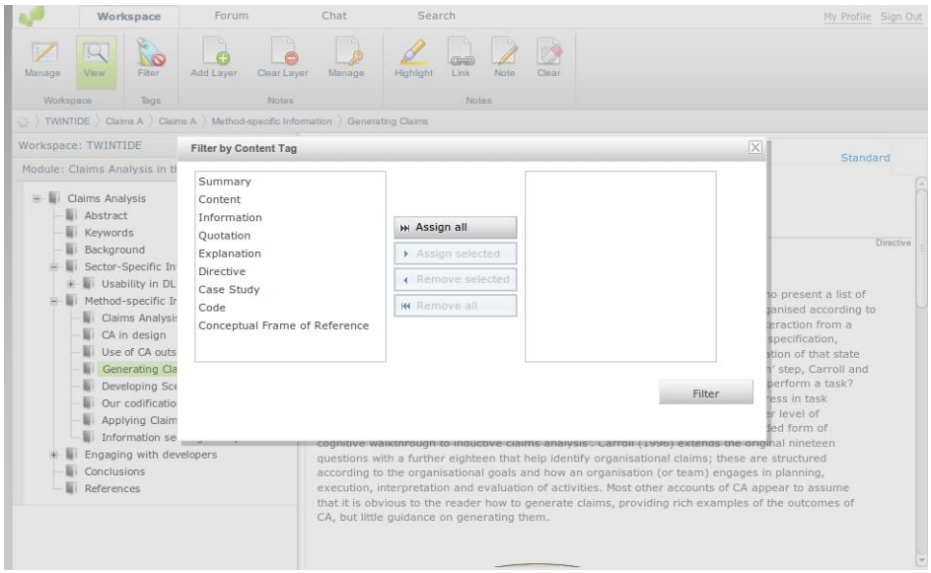



**Fig. 5.** Filter relevant information using metadata

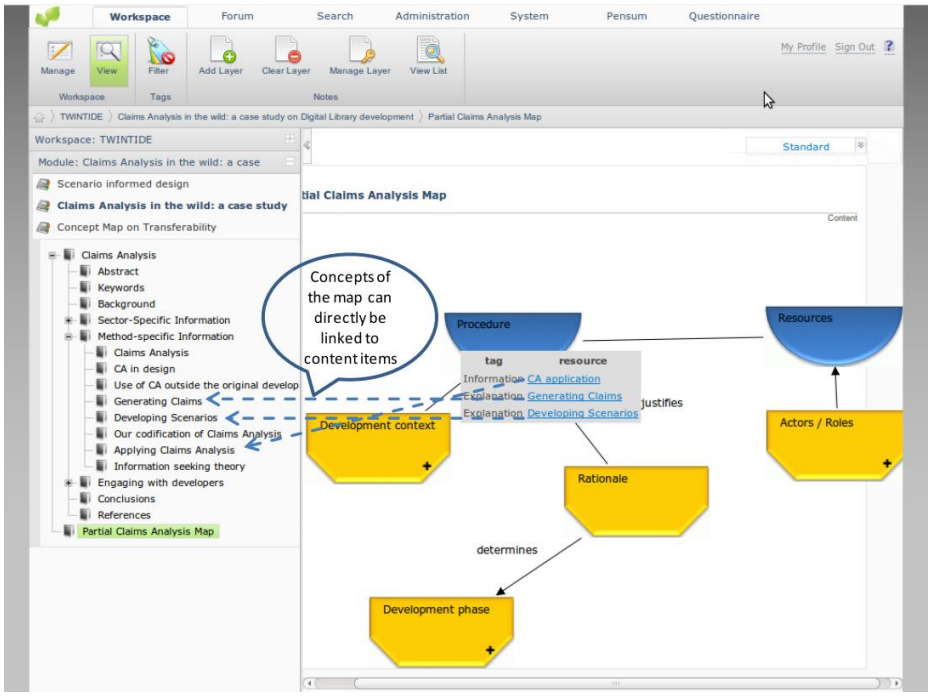



**Fig. 6.** Navigating transfer cases using the transfer map

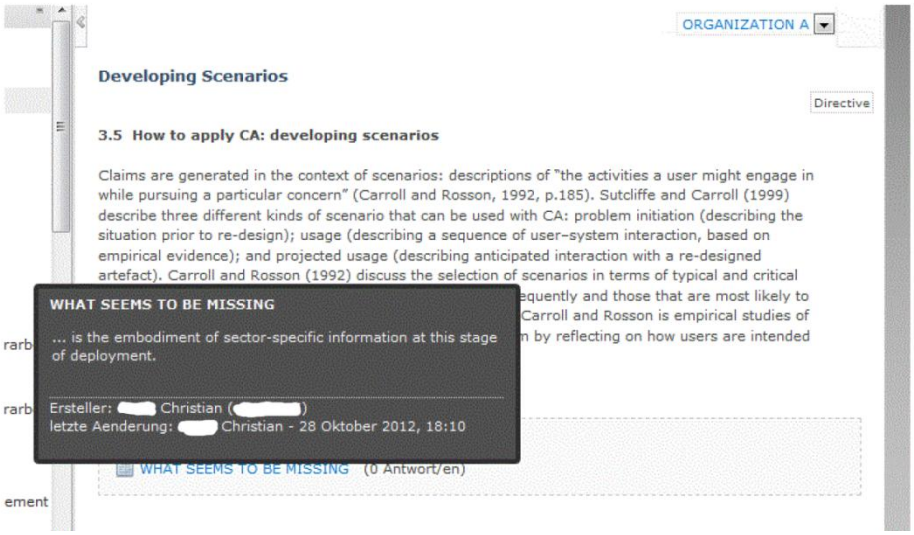



**Fig. 7.** Context-sensitive discussion of transfer cases

Within the Nymphaea platform overall 10 case studies have been restructured, enriched with metadata and provided for application and transfer relevant discussions and knowledge exchange.

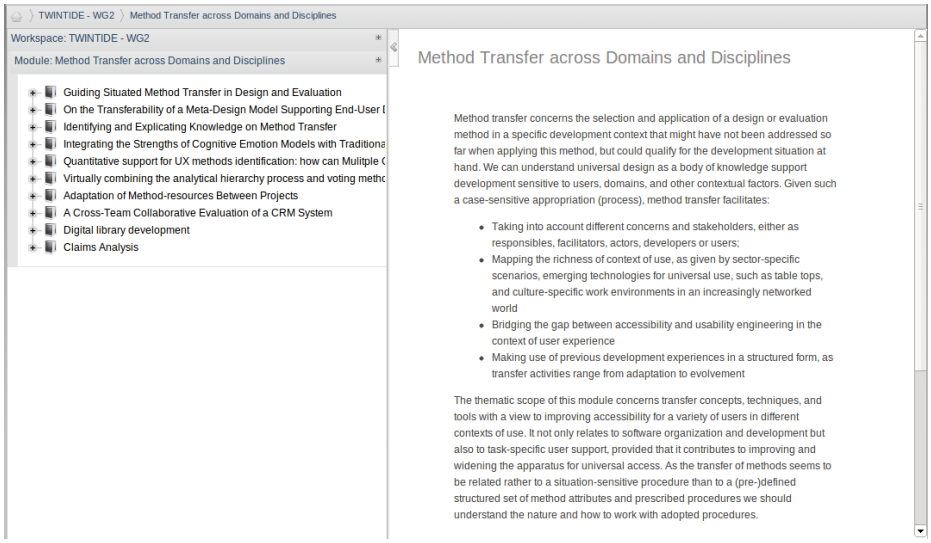



**Fig. 8.** Case Studies in Nymphaea considering
Method Transfer across Domains and Disciplines

The provided case studies in Nymphaea were published by:

- Wardlaw, J., Cox, A. L and Haklay, M. (2013): Adaptation of Method-resources Between Projects: A Case Study From a Dynamic and Complex Work Domain. CHI 2013 Workshop paper.
- Sikorski, M. (2013): A Cross-Team Collaborative Evaluation of a CRM System. CHI 2013 Workshop paper.
- Blandford, A., Keith, S., Butterworth, R., Fields, B., Furniss, D. (2007). Disrupting digital library development with scenario informed design. Interacting with Computers 19(1), 70-82 doi:10.1016/j.intcom.2006.07.003. Author URL http://eprints.ucl.ac.uk/5112
- Blandford, A., Keith, S., Fields, B. (2006). Claims analysis "in the wild": a case study on digital library development. International Journal of Human-Computer Interaction 21(2), 197-218 doi:10.1207/s15327590ijhc2102_5. Author URL http://eprints.ucl.ac.uk/5124
- Cronholm, S., Neubauer, M. Stary, C. (2013): Guiding Situated Method Transfer in Design and Evaluation: Exploring Concepts, Activities and Process. To be published in: Special UAIS Issue on 'Method Transfer across Domains and Disciplines: Enriching Universal-Access Development'. Springer.
- Hvannberg, E.T. (2013): Identifying and Explicating Knowledge on Method Transfer: A Sectorial System of Innovation Approach. To be published in: Special UAIS Issue on 'Method Transfer across Domains and Disciplines: Enriching Universal-Access Development'. Springer.
- Springett, M., Law, E., Coulson, M. (2013): Integrating the Strengths of Cognitive Emotion Models with Traditional HCI Interaction Analysis Tools. To be published in: Special UAIS Issue on 'Method Transfer across Domains and Disciplines: Enriching Universal-Access Development'. Springer.
- Srdjevic, B., Pipan, M., Srdjevic, Z., Blagojevic, B., Zoranovic, T. (2013): Virtually combining the analytical hierarchy process and voting methods in order to make group decisions. To be published in: Special UAIS Issue on 'Method Transfer across Domains and Disciplines: Enriching Universal-Access Development'. Springer.
- Melo, P., Jorge, L. (2013): Quantitative support for UX methods identification: how can Multiple Criteria Decision Making help? To be published in: Special UAIS Issue on 'Method Transfer across Domains and Disciplines: Enriching Universal-Access Development'. Springer.
- Ardito, C., Buonu, P., Francesca Costabile, M. et al. (2013): Meta-design Model Transfer for End-User Development. To be published in: Special UAIS Issue on 'Method Transfer across Domains and Disciplines: Enriching Universal-Access Development'. Springer.

In the course of the case study work the process perspective on transferring methods as well as the situation-specific handling of method attributes got evident. To investigate the handling of method attributes a so-called 'practical clean-up task' was performed in a subsequent meeting. The task within the meeting was defined as follows:

- How would you handle the key attributes of methods when transferring methods across sectors?

  Use document ‚Key-Attribute-List.docx‘

  - Groups of attributes
    - WHY, WHAT, HOW, WHAT ELSE
  - Detailed attribute descriptions
  - Resource type relationships
- Construct a process model in a
  - (diagrammatic) notation you feel confident with
- Provide rationale for each activity
  - Achievement in each step

The results showed different process models on the usage of the gained method attributes. They indicate different processes using different key attributes depending on the method chosen and the given situation. Fig. 9 illustrates one example that shows the consideration of Why, What, How aspects before a decision for a certain method is taken and the method chosen is applied. Within the consideration of Why, What and How aspects, sector-specific and method-specific activities should be considered.

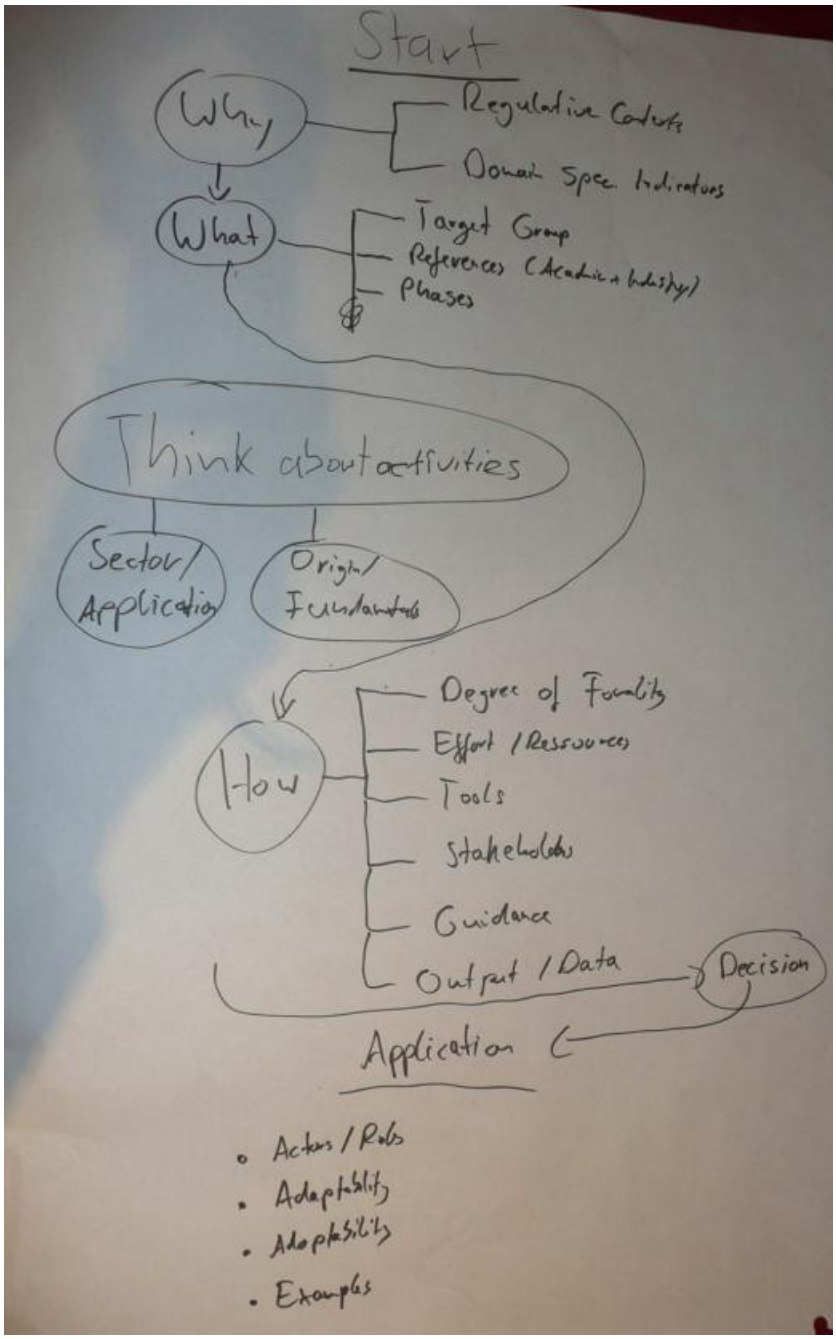



**Fig. 9.** Handling method attributes in transfer situations

The process perspective was further investigated in discussions on process models behind method selection approaches (e.g. Multi-criteria Decision Making) and tools such as Usability Planner (cf. Ferre, Bevan & Escobar, 2010; Fig. 10) and UCDtoolbox (cf. http://ucdtoolbox.com/). Results will be reported in the Springer Journal 'Universal Access to Information Society' within a special issue on 'Method Transfer across Domains and Disciplines: Enriching Universal-Access Development'.

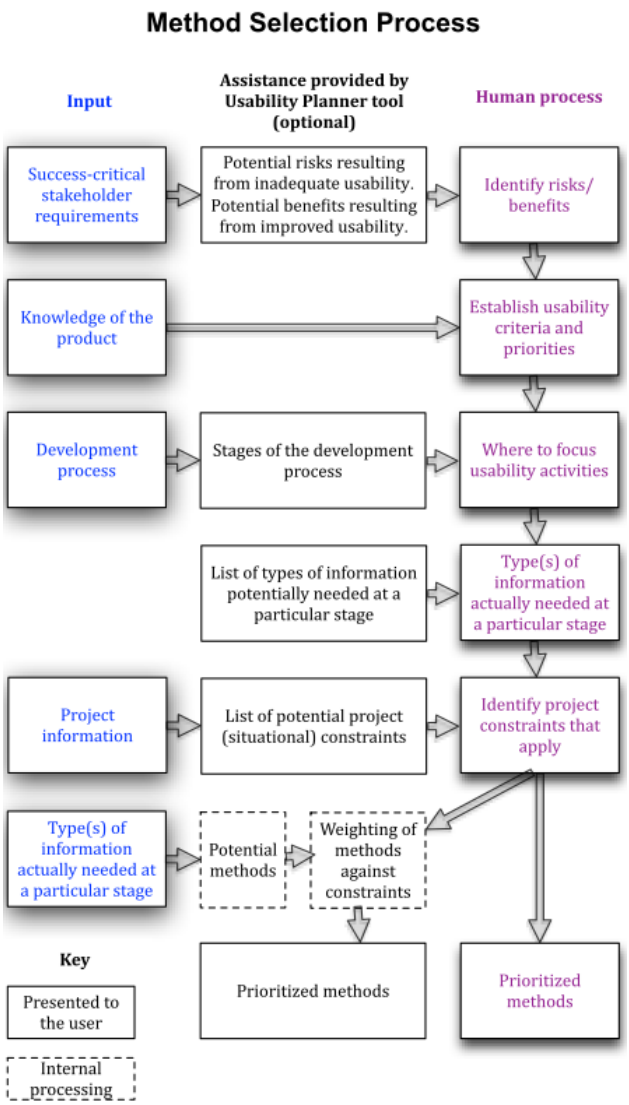



**Fig. 10.** Usability Planner – Method Selection Process by Nigel Bevan 18-Dec-12

# 3 Results and Discussion

The objective of WG2 was the identification and analysis of sector-unique design and evaluation methodologies in terms of their underlying conceptual models, practical protocols, and transferability to other sectors. The objective has been addressed by performing three main tasks using a method-mix described in section 2 'Approaching Transferability'. Subsequently, relationships between the three tasks and the methods used to elicit, represent and negotiated expert knowledge on method transfer are depicted in Fig. 11.

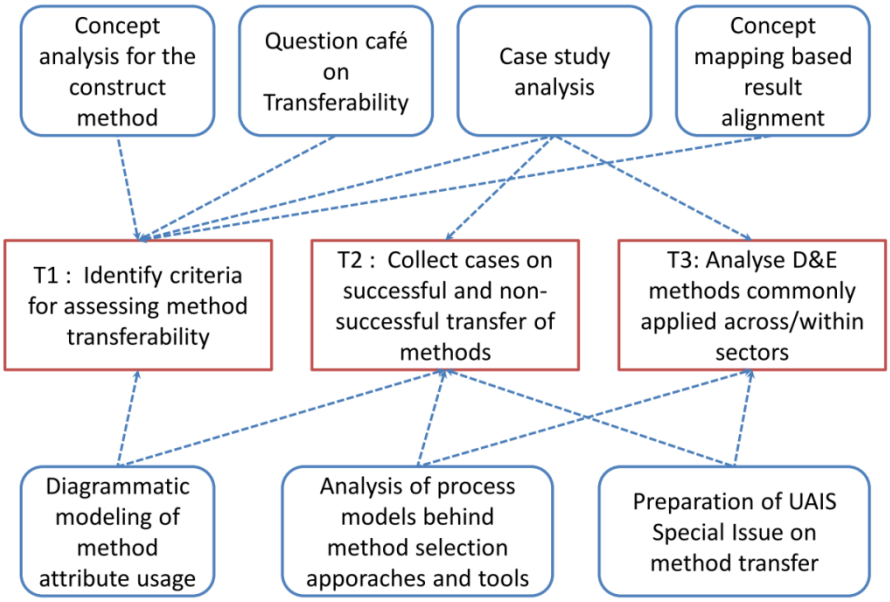



**Fig. 11.** WG2 methods used to perform project tasks

Overall, the objective and tasks of WG2 have successfully been reached. Detailed results of WG2 are provided within the TwinTide Wiki, the Nymphaea platform and in the upcoming Springer UAIS Special Issue on 'Method Transfer across Domains and Disciplines: Enriching Universal-Access Development'. In the following, selected results considering method transfer are presented.

Method transfer concerns the selection and application of a design or evaluation method in a specific development context that might have not been addressed so far when applying this method, but could qualify for the development situation at hand. We can understand universal design as a body of knowledge support development sensitive to users, domains, and other contextual factors. Given such a case-sensitive appropriation (process), method transfer facilitates:

- Taking into account different concerns and stakeholders, either as responsibles, facilitators, actors, developers or users;
- Mapping the richness of context of use, as given by sector-specific scenarios, emerging technologies for universal use, such as table tops, and culture-specific work environments in an increasingly networked world
- Bridging the gap between accessibility and usability engineering in the context of user experience

As the transfer of methods seems to be related rather to a situation-sensitive procedure than to a (pre-)defined structured set of method attributes and prescribed procedures we should understand the nature and how to work with adopted procedures. WG2 intention was to structure and foster the discussion about concepts and experiences when transferring methods across domains or sectors. Latest research work on the design and implementation of method transfer approaches between different application domains, scenarios of use, contexts, and sectors will be published in the UAIS Special Issue on 'Method Transfer across Domains and Disciplines'

The contribution of Stefan Cronholm et al. introduces the concept of situated method transfer in design and evaluation. Stefan Cronholm et al. explore transfer activities and develop a process structure aiming towards behavioral guidance when knowledge about method application is transferred from an experienced or documented development situation to the current one. The authors reflect the variety of methods for designing and evaluating interactive socio-technical systems. Taking into account their specifity with respect to development purposes and the observed diversification of interactive systems methods in an effective and efficient way has become decisive for development.

The authors propose a situation framework for learning from method applications in the context of usability of accessibility cases. It lays ground to informed procedures for using existing knowledge on the rationale of method application, domain-specific factors, related assumptions and constraints. Fig. 12 shows the frame of reference that can be used for capturing the essence of transfer situations. It provides information on the situation of applying a certain method, i.e. the situation before transfer (left side), and on the situation affected by this knowledge, i.e. the situation after transfer (right side). Each situation is detailed both, in terms of the concerned technique or method, and its characteristic (set of) items when in use, constituting the body of knowledge of a method or technique. These items can either be structural, such as roles, behavioral, such as a sequence of step to be followed when applying a method. One of the items should refer to sector or domain in each situation. The transfer itself is described by the type of activities that are set when moving from one development situation to another.

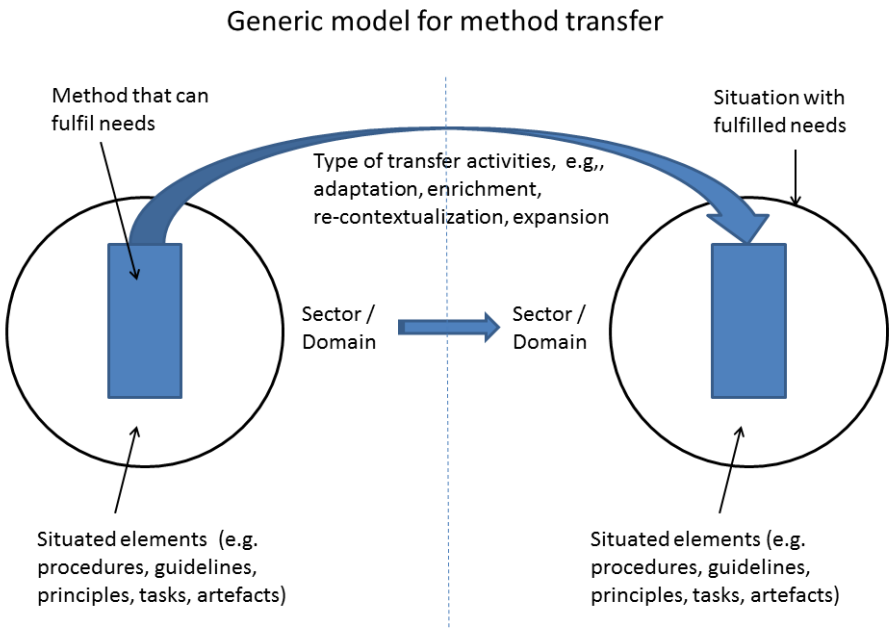



**Fig. 12.** Frame of reference describing situated transfer of methods

Cronholm et al. introduce a generic procedure that can be used for informed decision making in development projects with respect to methods. Finally, the authors introduce some guidelines that should support developers when identifying transfer-relevant items with respect to development situations involving methodological knowledge.

In the subsequent contribution the transferability of a meta-design model supporting End-User Development is addressed. Carmelo Ardito et al. show the transfer of such a model recognizing several key elements describing development context. It is part of structuring previous development experiences in a systematic way. The authors reports on experiences of transferring a model aiming at support of designing systems by different stakeholders. They need not to be professional developers, but to actively contribute to develop a system. As shown in Fig. 13 participatory creativity was the key for evolving the Software Shaping Workshop (SSW) meta model. By reflecting original SSW situation several key items could be identified. They facilitated the evolvement towards the Hive Mind Space (HMS) model, both from a structural and behavioral perspective.

Several sectors have been involved: The SSW model has been applied to healthcare and gaming, whereas the HMS model has been utilized for energy management. Consequently, the evolvement has also enriched application domains not being addressed so far by the approach. Informed design has been enabled by the rich set of situational parameters gained from implementing the SSW model.

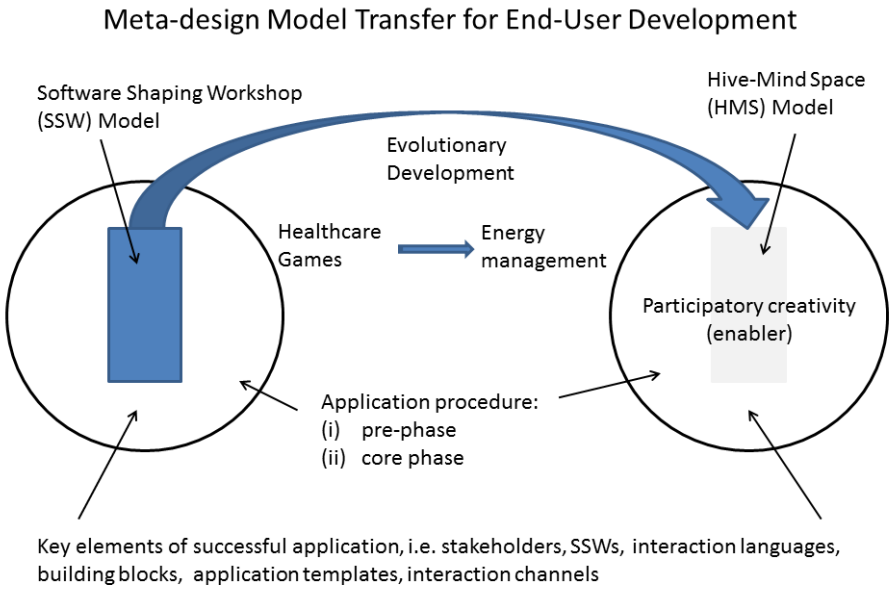



**Fig. 13.** Transfer through evolving a meta-model towards participatory creativity

In her contribution Ebba Hvannberg develops a sectorial system of innovation approach serving as framework when identifying and explicating knowledge on method transfer. The study reveals from two case studies the interaction within the sector of crisis management is as important as the interaction with the dimensions of methodological developments. Hence, a frame of reference for a sector facilitates handling the various aspects of method transfer, without restricting the type of methods being handled. It refers to knowledge and technologies, actors and networks, and institutions.

The proposed sectorial approach has been applied to evaluation methods, namely Heuristic Evaluation and User Testing. Both methods could be further developed in conformance with the facilities and capabilities of the researchers involved in the sectorial framework of innovation for crisis management. Fig. 14 reflects the situation affected by the transfer, identifying the 2 techniques as concerned subjects, including the existing body of knowledge on the HCI community. It has been embodied into each component of the innovation framework for crisis management, tuning it to domain-specific application.

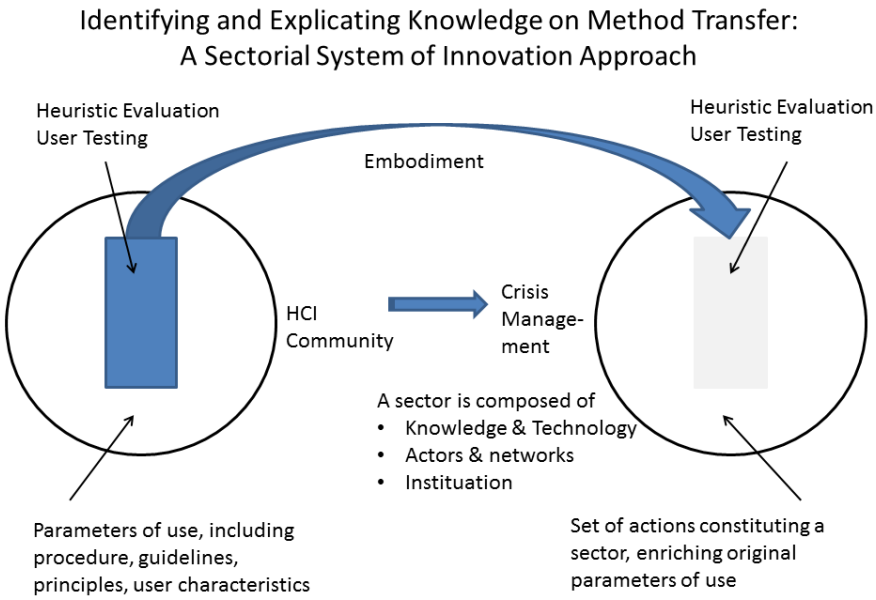



**Fig. 14.** Utilizing a sectorial system of innovation for embodying Usability Testing and Heuristic Evaluation

Mark Springett et al., in their contribution, deal with integrating the strengths of cognitive emotion models with traditional HCI interaction analysis tools. Their aim is to transfer the strengths of interaction models to analysis of affect critical systems. They start out identifying key concepts from cognitive models of emotion and cognitive models of interaction as given by the HCI literature. In particular, they look at concepts from Scherer's Appraisal model and Stimulation evaluation checks, and Coulson's framework of emotion contexts. This knowledge is transferred by integrating it into Norman's cycle of display-based action, as it has been applied for interaction analysis in human-computer interaction – see also Fig. 15.

The approach takes into account the recent shift of emphasis to user experience, which requires enriching purely cognition-based approaches to design with affective experiences such as happiness, joy and surprise. The proposed intertwining of Norman's cognition-based Action Cycle with models based on appraisal theories should allow shifting the focus towards understanding the causes and effects of feelings arising from interacting with digital artifacts, as indicated by several case studies with the elderly and e-shoppers.

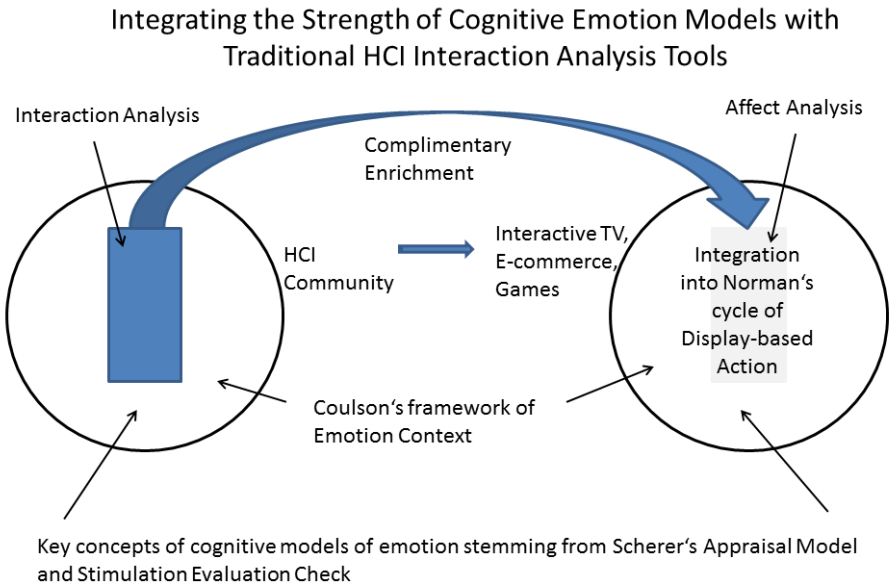



**Fig. 15.** Enriching interaction analysis by emotion model elements

The subsequent contribution also impacts design and evaluation, in particular when making decisions. In search for quantitative support for UX methods identification Melo et al. inspect multiple criteria decision making. They develop an application scenario on how to apply quantitative Multiple Criteria Decision Making (MCDM) approaches to certain aspects of User Experience (UX) design and evaluation. They target the identification of relevant methods in order to proceed with development activities.

In order to reflect strengths and weaknesses of methods, not only the context of use, i.e. the situation needs to be identified, bit also characteristic items of the methods need to be specified. Consequently, the proposed MCDM procedure starts out with consensus development on a set of parameters relevant to the methods and the development situation – see also Fig. 16. Then, a model for quantitatively matching can be applied corresponding to the transfer process. MCDM methods can be applied to support the informed selection of (UX) methods. Transfer in that context can be considered as configuration of method sets for a particular development context.

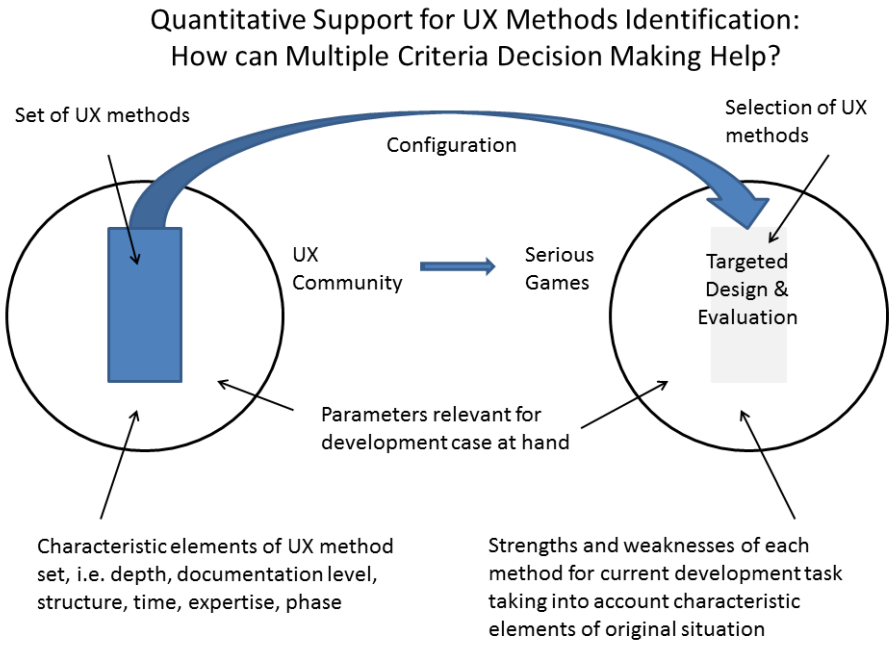



**Fig. 16.** Informed method set configuration

Multi-criteria decision making (MCDM) is also addressed by Srdjevic et al. when proposing a combining the Analytic Hierarchy Process and Social Choice Theory (SCT). The authors support group decision making in the course of design. MCDM is represented by the utility approach Analytic Hierarchy Process (AHP), as it can be considered most popular in group-decision support. SCT enriches the AHP with a voting system. It can efficiently be combined with AHP in various group-decision contexts.

Group decision-making processes can occur in the field of human-computer interaction using only AHP either seeking consensus or merely aggregating preferences. As the discourse seems to be important due to the variety of decision makers for universally accessible systems, SCT systems, both preferential and non-preferential voting systems need to be integrated into decision making. Hence, the authors propose to apply AHP in the first stage, in order to obtain weights of the alternatives and, in the second stage. Then AHP's cardinal information is interpreted as ordinal and used for SCT voting.

The case studies reported concern the utility of sources of information for search tasks. The results reveal plausible decisions, as the decision makers were expected to approve according to their (academic) background.

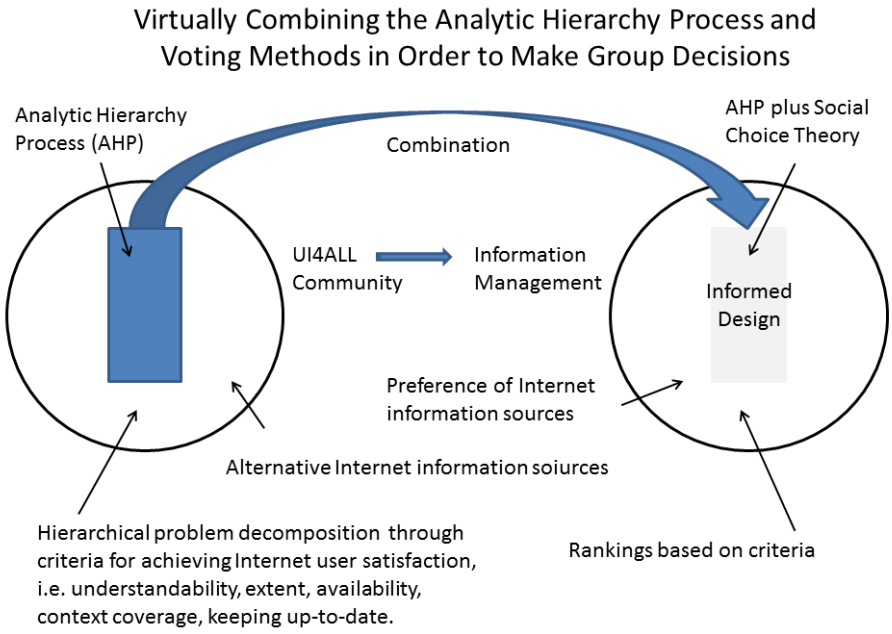



**Fig. 17.** Combining Analytical Hierarchical Process with Social Theory for informed design decisions

Overall, the contributions indicate that transparent method characteristics and context-of-use parameters referring to the development situation can facilitate the transfer of methodological knowledge. It is the reflected situation-sensitive match of parameters leading to informed design or evaluation decisions.

# 4 Impacts Achieved

## 4.1 List of publications


1. Cronholm, S., Neubauer, M. Stary, C. (2013): Guiding Situated Method Transfer in Design and Evaluation: Exploring Concepts, Activities and Process. To be published in: Special UAIS Issue on 'Method Transfer across Domains and Disciplines: Enriching Universal-Access Development'. Springer.
2. Hvannberg, E.T. (2013): Identifying and Explicating Knowledge on Method Transfer: A Sectorial System of Innovation Approach. To be published in: Special UAIS Issue on 'Method Transfer across Domains and Disciplines: Enriching Universal-Access Development'. Springer.
3. Springett, M., Law, E., Coulson, M. (2013): Integrating the Strengths of Cognitive Emotion Models with Traditional HCI Interaction Analysis Tools. To be published in: Special UAIS Issue on 'Method Transfer across Domains and Disciplines: Enriching Universal-Access Development'. Springer.
4. Srdjevic, B., Pipan, M., Srdjevic, Z., Blagojevic, B., Zoranovic, T. (2013): Virtually combining the analytical hierarchy process and voting methods in order to make group decisions. To be published in: Special UAIS Issue on 'Method Transfer across Domains and Disciplines: Enriching Universal-Access Development'. Springer.
5. Melo, P., Jorge, L. (2013): Quantitative support for UX methods identification: how can Multiple Criteria Decision Making help? To be published in: Special UAIS Issue on 'Method Transfer across Domains and Disciplines: Enriching Universal-Access Development'. Springer.
6. Ardito, C., Buonu, P., Francesca Costabile, M. et al. (2013): Meta-design Model Transfer for End-User Development. To be published in: Special UAIS Issue on 'Method Transfer across Domains and Disciplines: Enriching Universal-Access Development'. Springer.
7. Neubauer, M., Oppl, S., Stary, C., & Weichhart, G. (2013). Facilitating Knowledge Transfer in IANES-A Transactive Memory Approach. In Innovation through Knowledge Transfer 2012 (pp. 39-50). Springer Berlin Heidelberg.
8. Stary, C. (2012). Growing through learning–Essential technology enablers. Journal of Organisation Transformation & Social Change, 9(1), 63-74.
9. Fleischmann, A., Schmidt, W., & Stary, C. (2013). Open S-BPM= Open Innovation. In S-BPM ONE-Running Processes (pp. 295-320). Springer Berlin Heidelberg.
10. Fleischmann, A., & Stary, C. (2012). Whom to talk to? A stakeholder perspective on business process development. Universal Access in the Information Society, 11(2), 125-150.
11. Fleischmann, A., Schmidt, W., Stary, Ch., Obermeier, St., Börger, E. (2012) Subject-oriented Business Process Management, Springer Berlin Heidelberg.

### 4.2 List of outreach activities

***Outreach to South Africa:*** In the context of the ICKM (Int. Conf. on Knowledge Management) – University of Johannesburg, Prof. duToit, several research cafés have been hold w.r.t. method transfer, in particular, eliciting knowledge on acquisition and learning, involving practitioners and academics. The inputs have been taken up in the work of the Council on KM (www.ickm.net) and its members. Endorsement of KM programs and curricula is facilitated by TwinTide experiences and input to cross-disciplinary method transfer.

***Outreach to Macedonia:*** Establishment of bilaterally funded research-teaching collaboration with Prof. Saso Josimovski, Faculty of Economics, Skopje, Macedonia. Active research and lecture exchange with Matthias Neubauer through mutual visits, also including Martin Mihajlov. Fields of research: interactive elicitation of process knowledge. TwinTide formats and the framework for method transfer are applied in that context.

## 5 Conclusion and Future Work

Working Group 2 of the TwinTide project aimed to identify and analyze sector-unique design and evaluation methodologies in terms of their underlying conceptual models, practical protocols, and transferability to other sectors. This objective has been addressed by performing three main tasks using a method-mix described in section 2 ‘ supporting the elicitation, representation and negotiation of expert knowledge on method transfer. Overall, the working tasks have successfully been performed and the objective of WG2 has been reached. The results are intended to structure and foster the discussion about concepts and experiences when transferring methods across domains or sectors.

As a result method transfer is understood as the selection and application of a design or evaluation method in a specific development context that might have not been addressed so far when applying this method, but could qualify for the development situation at hand. Transparent method characteristics and context-of-use parameters referring to the development situation can facilitate the transfer of methodological knowledge. It is the reflected situation-sensitive match of parameters leading to informed design or evaluation decisions.

Future work should consider further validation of the gained knowledge within concrete transfer situations as well as the reporting on experiences within such situations. Nymphaea could be one means for the community to exchange knowledge on method transfer and build a common knowledge base for experts as well as practitioners.

# 6 References


1. Dalkir, K. (2005). Knowledge management in theory and practice. Butterworth-Heinemann.
2. Rawlinson, J. (1981). Creative thinking and brainstorming. Gower England.
3. Bowman, D., Gabbard, J., and Hix, D.: A Survey of Usability Evaluation in Virtual Environments: Classification and Comparison of Methods. Presence: Teleoperators and Virtual Environments, vol. 11, no. 4, 2002, pp. 404-424
4. Blandford, A., Keith, S., Butterworth, R., Fields, B., Furniss, D. (2007). Disrupting digital library development with scenario informed design. Interacting with Computers 19(1), 70-82 doi:10.1016/j.intcom.2006.07.003. Author URL http://eprints.ucl.ac.uk/5112
5. Blandford, A., Keith, S., Fields, B. (2006). Claims analysis "in the wild": a case study on digital library development. International Journal of Human-Computer Interaction 21(2), 197-218 doi:10.1207/s15327590ijhc2102_5. Author URL http://eprints.ucl.ac.uk/5124
6. Ferre, X., Bevan, N., Escobar, T.A. (2010): UCD Method Selection with Usability Planner. IN: Proceedings of NordiCHI 2010, October 16-20, 2010.
7. Fitzpatrick, R. : Strategies for Evaluating Software Usability, Department of Mathematics, Statistics and Computer Science, Dublin Institute of Technology, Kevin Street, Dublin 8, Ireland 1999
8. Introna, L. D. and Whitley, E. A.: Against method-ism Exploring the limits of method. Information Technology & People, Vol. 10 No. 1, 1997, pp. 31-45. MCB University Press.
9. Steves, M., Morse, E., Gutwin, C., Greenberg, S. 2001. A Comparison of Usage Evaluation and Inspection Methods for Assessing Groupware Usability. In Proceedings of the ACM Conference on Supporting Group Work, 125-134
10. Szulanski, G.: Exploring Internal Stickiness: Impediments to the Transfer of Best Practice Within the Firm. Strategic Management Journal, Vol. 17 (1996), pp. 27-43, doi:10.2307/2486989

# 7 Appendices

## 7.1 Concept Analysis for the construct 'method'

**Abbreviations**
**Tbd** to be discussed

**Cluster Why.**

**Rationale: objectives, goal**
* Examples: improvement of work conditions, identification of short-comings, providing evidence of success, find usability problems,
* Non-Examples: monitoring of work conditions (since it is a task)
**Regulative context**
* Examples: EU directive on Human-Machine Interaction, accessability standards (?)
* Non-Examples:
**Domain specific indicators**
* Examples: is a standard procedure in mechanical engineering
* Non-Examples:

**Cluster What**

**Origin, fundamentals**
* Examples: ACT*, Grounded Theory, ethnography
* Non-Examples:
**Involved disciplines and areas**
**G4 – theoretical underpinnings (tbd)**
* Examples: occupation psychology, pedagogy,
* Non-Examples:
**Author(s)**
* Examples:
* Non-Examples:
**Name of method**
**G4 – impact of name – important in terms of marketing it in respective area**
* Examples:
* Non-Examples:
**Scope: Design, evaluation, analysis to implementation**
* Examples:
* Non-Examples:
**Depth**
* Examples: work task level
* Non-Examples:
**Target group(s)**
* Examples: quality management, designers, developers

* Non-Examples:
**Sector / Application area**
* Examples: e-Learning, e-Government, gaming
* Non-Examples:
**References (sources of information)**
* Examples:
* Non-Examples:
**Theoretical foundation/background** How does this differ from "origin" above?
* Examples:
* Non-Examples:
**Concepts / Ontology [What are the things a method talks about? E.g. GOMS has a strong ontology compared to heuristic eval.]**
**[tbd] – is it a key attribute?**
* Examples: situation awareness in control rooms [SAGAT],
* Non-Examples:

**Cluster How**

**Actors or roles**
* Examples: domain expert, end-user
* Non-Examples: demographic description of people
**Procedure: Steps to follow**
* Examples: steps to take in a cognitive walkthrough, creating + using UML diagrams
* Non-Examples: UML Diagrams themselves
**Degree of formality: formal/semi-formal/informal**
**G1: - of following the procedure**
* Examples: GOMS (formal), heuristic Eval. (informal),
* Non-Examples: think aloud (not a characteristic)
Tbd – to be disc. [key attrib?]
**Using What: Data, evidence**
* Examples: audit templates, data 4 key stroke modeling
* Non-Examples: anything you may need 4 cultural probes
**Constraints / Context of use [tbd – any method without constraints?]**
* Examples: in the course of certification (of what?)
* Non-Examples:
**Effort to be spent**
**G1 – changed to resources needed**
* Examples: 30 minutes, time req., number of users, costs,
* Non-Examples: overall project time
**Elapsed time to complete a method**
**Tools**
* Examples: mindmapping [tbd], standard questionnaire[SUMI]
* Non-Examples: database behind [SUMI]
**Tool support**
* Examples: database behind [SUMI]

* Non-Examples: usage of SPSS

**Degree/type of involvement of stakeholders**

**Researcher presence/involvement**

* Examples: absence of researcher in case of an online survey, empathy, effort needed in PDP

* Non-Examples: amount/number of users in user testing

**Structuredness/guidance**

* Examples:

* Non-Examples:

**Time/resources required**

* Examples:

* Non-Examples:

**Examples of application**

**G4: rename in "appropriate uses"**

* Examples:

* Non-Examples:

**Composability (other methods which are complementary)**

* Examples:

* Non-Examples:

**Adaptability**

* Examples:

* Non-Examples:

**Adoptability (known variants of the method)**

* Examples:

* Non-Examples:

**Context**

* Examples: field method, lab method, online method

* Non-Examples:

**Output / type of Data collected**

**[tbd]**

* Examples **(quant./qual.)**: screen-recording, log-file, video recording, audio recording

* Non-Examples: -

**Quantification measures available?**

* Examples:

* Non-Examples:

**Representations generated/used**

* Examples: usage statistics, persona, wireframe (?)

* Non-Examples: histograms, bar charts (representation description should refer to content)

**Required expertise**

* Examples:

* Non-Examples:

**Cluster What Else.**

**Practical Experiences**
* Examples:
* Non-Examples:
**Empirical evidence**
* Examples:
* Non-Examples:
**Learnability**
* Examples:
* Non-Examples:
**Efficiency / effectiveness**
* Examples:
* Non-Examples:
**Available/support literature**
* Examples:
* Non-Examples:
**Validity**
* Examples:
* Non-Examples:
**Replicability**
* Examples:
* Non-Examples:
**Reliability**
* Examples:
* Non-Examples:

**Accuracy (not sure whether this adds to existing attributes validity and reliability)**
* Examples:
* Non-Examples:

**Ethical considerations or constraints**
* Examples:
* Non-Examples:
**Limits**
* Examples:
*Non-Examples:

## 7.2 Towards Transferability: Results of the Question Café in Reykjavik

**Project specific factors**

- What are the constraints/differences coming from the organizational culture?
  - Because every organization is situated in a broader socio-economic environment that determines what is expected or valued
- Is trust important?
  - Because Design effects trust of the users
- What kind of development process is adopted?
  - Because methods have to fit type and stage of process
- Are there laws or restrictions for being allowed or having to use specific methods?
  - Because this limits the pool of methods you can use (e.g. prescribed methods in medical field)
- Is the application safety or business critical?
  - Because this will prioritize using different types of methods
- Are the users mostly first time users, regular users or both?
  - Because this should inform design and evaluation
- Is there one specific application context (professional system) or many application contexts: type of users, range of tasks and environments
  - Because this should inform design and evaluation
- On which types of platforms (e.g. desktop, mobile,…) you plan to use? How many?
  - Because this information is important to understand how many technologies to consider
- Which app context are we interested in?
  - There are at least 2
  - We may learn about both (original context, Transfer context)

***Similar questions***

- What is more important – accurate, efficient performance or engagement and user experience?
- What are the main concerns of the domain (e.g. experience, safety-realibility)?
  - Because this will determine what kinds of method are most applicable
- Which resources are available?
  - Because, different methods require different resources
  - Because the availability of resources determines how and which methods can be used

**Method specific factors.**

- Which development processes does the method fit with?
  - Because target sector/domain might use different development process
- Which kind of stakeholders you need for using a method?
  - Because you need to check if you have them when transferring to another sector
- How malleable / adaptable is the method?
  - Because this determines how easily it can be adapted to new situations
- What are results from the method?
  - Because the format of the results may affect how they can be exploited in different sectors
- What resources are needed to apply the method?
  - Because it is necessary to match them to the resources of the domain / sector
- What is the (purpose / scope) of the method?
  - Because we need to know whether the method can address the needs of a particular domain / sector
- What artifacts does the method require?
  - Examples of artefacts are list of heuristics, system prototypes, specific forms needed by the method
- How does a method fit with the current development culture?
  - Because the method has to be accepted by the teams and results have to be accepted by stakeholders
- How reliable is the method?
  - Because this is important for acceptance in a new domain
- What expertise is needed to carry out the method?
  - Because the needed expertise might impose some training etc. issues

**Transfer specific factors.**

- What “values are embodied in or communicated through a method?
  - Because transfer depends on the ability to promote a method in a new context
- Are there related methods in the new domain?
  - Because the transferred method should take into account what is already available
- What added value would the transferred method provide in the new domain?
  - Because unless there is added value, the method will not be used
- Are there cultural factors that determine particular methods?
  - Because these influence whether or not a method can be transferred effectively to a new context
- What key attributes can be changed?
  - To determine which parts of a method are adoptable
- What key-attributes are fixed?
  - To determine which parts of a method can not be changed

- What key attributes should be changed?
  - Which parts of a method should be changed so it can work in a new domain
- When is it an adapted method vs. a new method?
  - At which moment is a new method too different from the original method so you actually have created a new method?
- How much experience/knowledge is there of using the method successfully in current domain?
  - Because this will facilitate our understanding of successfully transferring a method
- What are domain specific factors? In what respect do the domains differ (e.g. type of products, development process, target users, environment,…)
  - To determine which factors should be taken into account when adapting a method
- How much time and expertise is available for the transfer?
  - Because having more time available would allow the company to adapt to the requirements of the new methods
  - Because more expertise would facilitate (or be required) for implementing the method
- Once transferred, can the method be used? (expertise, resources)
  - Because resources to transfer != resources to use

### 7.3 Cases on method application and method transfer

The cases are provided within the nymphaea platform. The platform is available at:

- https://nymphaea.ce.jku.at
  - Username: twintide
  - Password: twintide

# TwinTide WG3 Report

## Interplay between Design and Evaluation, Quality Models and Standards

**Ebba Þóra Hvannberg**

University of Iceland

With contributions from:
Stephan Weibelzahl, *National College of Ireland*
Costin Pribeanu, *ICI Bucharest, Romania*
Panayiotis Zaphiris, *Cyprus University of Technology*

# WG3: Interplay between Design and Evaluation, Quality Models and Standards

Ebba Þóra Hvannberg

Department of Faculty of Industrial Engineering, Mechanical Engineering and Computer Science, University of Iceland
ebba@hi.is

**Abstract.** This report is a summary of the activities carried out within the TwinTide project, working group 3. It describes the objectives, approaches, results, impacts and future work.

## 1 Objectives

### 1.1 Initial objectives

This section describes the objectives as they were set forward in TwinTide's proposal. The main goal of this activity was to identify and critically review the quality attributes and software standards that are currently prioritized in design and evaluation processes in different sectors. Human-centred approaches to interactive software development have emphasised first usability and now user experience. Some HCI researchers and practitioners propose specific qualities of successful interaction, such as Jeff Raskin's automaticity (Raskin, 2000). However, other qualities are given prominence in other software practices, for example, creativity in digital media. In high dependency systems, some combination of safety, security and reliability is required, with these qualities thus prominent in software engineering practice and research in relevant application domains. More recently, an emerging perspective within software engineering has stressed the quality of evolvability as being crucial to adapt alongside dynamic domain-specific models, user experiences, and organizational processes. The challenge in the development of interactive software is to balance these qualities, whether they relate to human usage, application performance or the software development process itself. Emphases on, and tradeoffs between, qualities vary with design purpose. At the most abstract level, balances and trade-offs here can be related to the "Quality in Software, Interaction and Value" (Q-SIV) framework, which is one of the fruitful outcomes of COST294 (Law, Hvannberg, & Cockton, 2008). The Q-SIV framework addresses quality in software by looking at how using development tools can enhance usability or other qualities such as security of a system, and how methods and models can be integrated into the development process; to address quality in interaction, the Action's partners can apply theoretical frameworks on the nature of interactions and methodologies to evaluate

qualities such as usability, automaticity or delightfulness; to address quality in value, the Action's partners assess the impact that a system has in the real world, focusing on both increasing value for software development and on increasing value for users and other stakeholders.

The original focus of WG3 was on the development of the related software standards (e.g. ISO 9241 series, particularly ISO 9241-210 "Human-centred design processes for interactive systems" (DIS, 2010) and ISO 9241-230 "Human-centred design methods"), which have been revised and created in response to the shift in the field of HCI. In particular, the Action aims to contribute to the refinement of such standards. Thematic topics include how different sectors appropriate and incorporate standards in their products/services, and which quality factors (e.g. safety, usability, reliability, evolvability) individual sectors prioritize and how they are evaluated.

The following tasks were stated in TwinTide's proposal.

WG3a. To identify what software qualities (e.g., trust, creativity, automaticity, security, safety, sociability, usability, reliability, evolvability) are prioritized and realized in the process of software development in different sectors/disciplines;

WG3b. To identify which standards address the emerging software qualities such as trust, sociability, and creativity and what their impacts on the real practice are;

WG3c. To understand how iterative design-evaluation-redesign feedback cycles operate for computing systems in different sectors;

For the task WG3a, it is relevant to collect quality models and quality management documents of individual R&D projects. For a project in a particular sector, the D&E methods applied may (not) enable the attainment of the software qualities prioritized. It is relevant to study the relationships between these two aspects. The task WG3b will effectively be dealt with given the strong representation of the Action's partners in the standardization bodies. WG3c is a more challenging task as it is necessary to track the development of the computing system of interest. Alternatively, interviewing the development team concerned may yield some relevant data, albeit the validity of the results may somewhat compromise. Specific attention will be granted to emergent D&E methods under the rubric "reflective design" where qualitative approaches are applied, given its highly fragmented status and broad cross-sectorial relevance.

The expected outcomes of the working group were a generic quality model for computing systems in a specific sector may be derived from the outcomes of the task WG3a. It was expected that the results of the task WG3b might lead to concrete proposals to the standardization bodies on the refinement of defining and operationalizing certain quality attributes. WG3c was expected to contribute to the ongoing research on downstream utility of D&E methods

### 1.2 Modified objectives

At the kick-off meeting and based on members' interests there was less emphasis on standards and most emphasis on researching software qualities across different sectors. Hence, the following objectives were addressed WG3a and WG3c.

WG3a. To identify what software qualities (e.g., trust, creativity, automaticity, security, safety, sociability, usability, reliability, evolvability) (related to standards, e.g., ISO 9241, 27477))

- are differentiated by sector-dependent or sector-independent criteria and describe connections
- vary over lifecycle
- are traded-off against quality attributes
- are prioritized and realized in the process of systems development in different sectors/disciplines

WG3c. Understand how iterative design-evaluation-redesign feedback cycles operate for

- computing systems in **different sectors**
- relative to specific evaluation criteria: social, economic, technical etc.
- traceability through cycle

## 2 Approaches

### 2.1 Workshops and meetings

WG3 activities were discussed and presented at eight TwinTide meetings. One internal workshop was dedicated to the research activities of the working group in Londin in 2012. Further, its status was briefly reviewed at two other internal workshops in Linz, 2011 and in Delft 2012. The results of the internal workshops in London motivated the open workshop at CHI'13: Made for sharing: HCI Stories of Transfer, Triumph and Tragedy (Law, Hvannberg, Vermeeren, Cockton, & Jokela, 2013).

**Fig. 1.** Working on stories in London, 2012

## 2.2 Activities

The research methods used within the action were systematic literature reviews and expertise knowledge extraction.

**Table 1** shows a list of meetings and activities carried out. After each of the activities we have labelled them with the tasks a or c according to the two objectives of the working group.

**Table 1.** List of meetings and activities

| Meetings and year | Activities |
|---|---|
| Brussels Meeting, 2009 | Revision of research tasks |
| Newcastle Meeting, 2010 | Members‘ interests and research mapped |
| Limassol Meeting, 2011 | • Sectoral System of Innovation Framework introduced (c )<br>• Questions on quality attributes of sectors identified (a)<br>• Questions on interplay between design and evaluation identified (c) |
| Bertinoro Meeting, 2011 | • Quality attribute maps of relations between domains (c) |
| London Workshop, 2012 | • Literature analysis of studies involving quality attributes applied in an application domain or a sector while studying methods of evaluation or design (a, c) |

| | |
|---|---|
| | • Cross-check between quality attribute maps to literature (a)<br>• Criteria of domain specific attributes (a)<br>• Telling stories of transferring methods (c) |
| Skopje meeting, 2012 | • Consolidating quality maps for sectors (a)<br>• Prioritization of questions on qualities and interplay (a, c) |
| Copenhagen meeting, 2012 | • Telling two stories of method applications (c) |
| Coimbra meeting, 2013 | • Refining stories<br>• Comparing method resources of stories ( c )<br>• Comparing contextual factors of stories ( c ) |
| Tallinn meeting, 2013 | Preparation of final report<br>Analysing stories |
| Made for sharing: HCI Stories of Transfer, Triumph and Tragedy Workshop at CHI 2013 | Discussion on the case studies |

The research methods applied for each of the tasks are provided in more detail below, starting with the objective of analyzing quality attributes between different sectors.

### 2.3 Research method: Quality attributes in different sectors

This study comprised two parts, experts mapping quality attribute to sectors and a systematic literature review. First, we review the method for mapping quality attributes. The task comprised several steps:

- Quality attribute maps of relations between sectors creating a baseline quality map
- Cross-check between quality attribute maps to literature (a)
- Criteria of domain specific attributes (a)
- Consolidating quality maps for sectors (a)

The research method for the literature analysis was a systematic literature review (Kitchenham & Charters, 2007). In a call for contributions, members were invited to find and briefly analyse papers which discuss a method to evaluate or design user interfaces / human computer interactions with the explicit criteria of including a certain quality attribute. Papers which address certain application domain(s) should be selected. Papers, which show transfer of methods between sectors, should be given more weight. Thus, there are four essential criteria (a-d) and one desirable criterion (e) for the paper selection:

a. Address one or more quality attributes of software

b. Address an application domain or a sector
c. Address a method of evaluation or design
d. Be published in a conference or a journal
e. Optional: Illustrate, implicitly or explicitly, transfer of a method between sectors

Criteria for exclusion were given as papers:
I. That address quality attributes of methods only
II. That have little or no grounding in empirical data
III. That address comparison of methods
IV. Duplicate reports of the same study (please include the most complete one)

A brief analysis of the paper should include the following items that were provided in a template accessible on the web:
i. A full reference of the paper
ii. Affiliations of authors
iii. A short summary of the study, including main research questions and answers
iv. Quality evaluation of the paper, i.e. impact of the journal, citations.
v. For each of the criteria (a)-(c) and possibly (e), members were asked to describe briefly how the paper addresses the item with exact reference to page number(s) and paragraph(s).

At the meeting work participants worked in groups of 2-3 to review/check each others contributions.

### 2.4 Research method: Design and evaluation methods in different sectors

The following steps were carried out to learn more about design and evaluation methods in different sectors and transfer therein:

- Telling stories of transferring methods
- Telling two stories of method applications
- Refining stories
- Comparing method resources of stories
- Comparing contextual factors of stories

## 3 Results and discussions

### 3.1 WG3a Quality attributes

There were two activities carried out within this task, a mapping of quality attributes of individual sectors and literature analysis on quality attributes. The following two subsections describe the results of these efforts.

**Mapping of quality attributes of individual sectors.**

The results of the quality attributes of sectors mentioned by expert members were drawn as a map that can be shown in **Fig. 2.** Initial quality attribute map**Fig. 2**. It shows that members think that there specific quality attributes to only some sectors. Since members were only asked to name quality attributes that had common sectors, we do not know if they thought there were some quality attributes that were unique to a sector. Notably the quality attributes mentioned were the traditional ones of usability, such as efficiency, effectiveness and learnability, more recent ones such as sociability and trust, attributes typically linked to user experience such as pleasure, but also quality attributes such as integrity and task appropriateness that are not necessarily linked to human computer interaction, but are on quality of results (i.e. content).

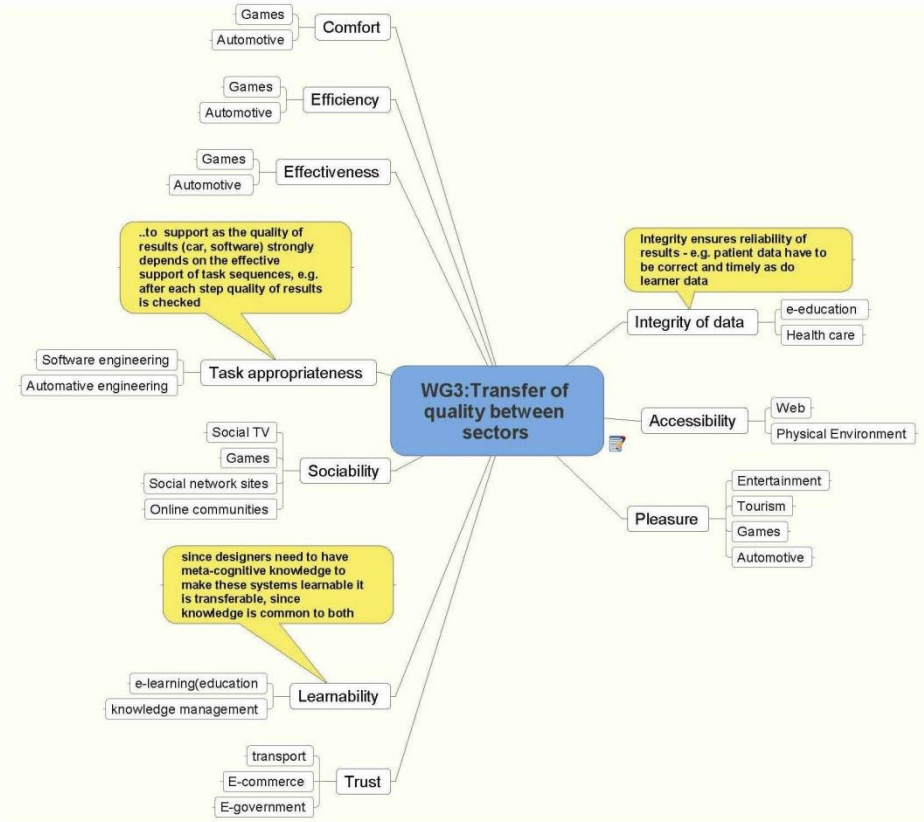



**Fig. 2.** Initial quality attribute map

Using the quality attribute in **Fig. 2** as a basis, six groups of 3-4 members were asked to extend it. Members were all experts and had all recently reviewed papers in the London workshop on quality attributes in different sectors. We will briefly summarise their findings below. Example map is provided in **Fig. 3**

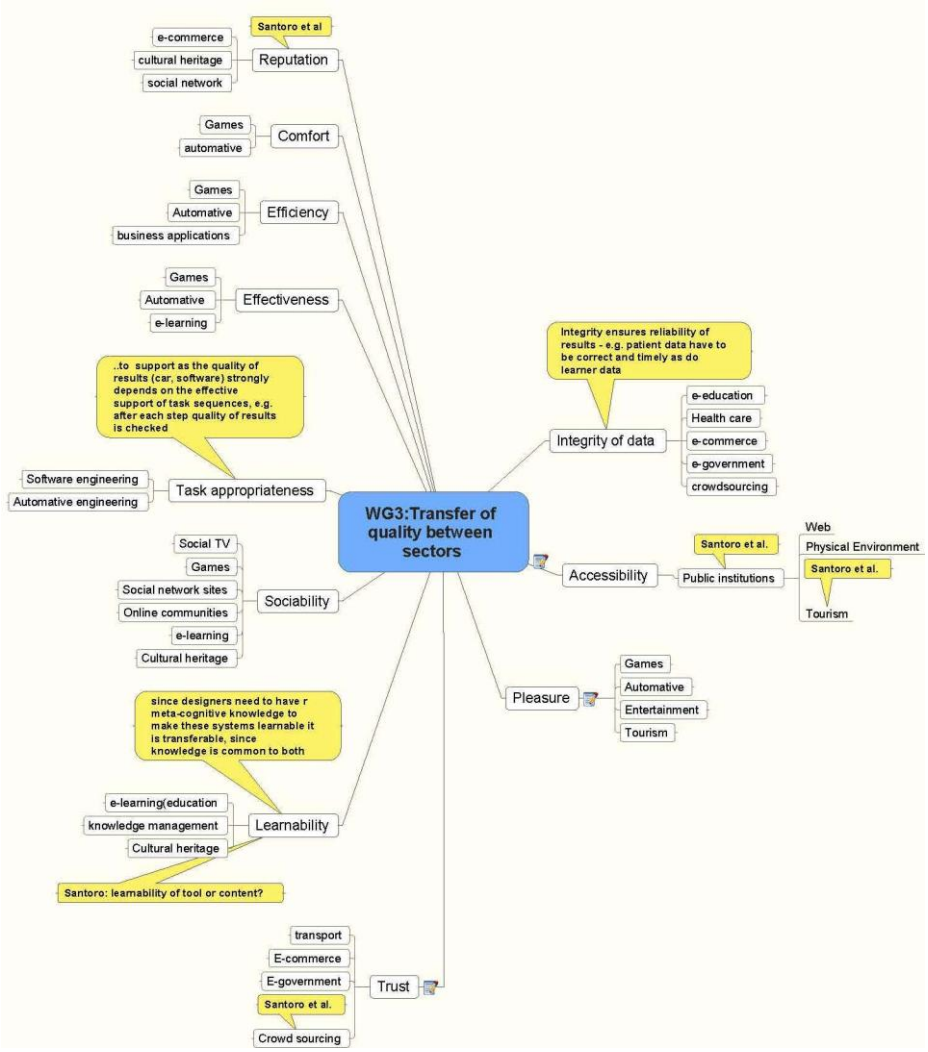



**Fig. 3.** Example mapping of quality attributes to sectors by experts (Group 2)

Group 1 added satisfaction and engagement without attaching it to any sector. The group added content quality (e.g. web content) and divided it into product quality, objective and subjective quality. Similarly, the group suggested usefulness as a quality attribute and noting its three perspectives, i.e. product, objective and subjective quality.

Group 2 added reputation as a quality attribute and linked it to the sectors e-commerce, cultural heritage and social network. They added cultural heritage to the learnability attribute and crowd sourcing to trust and integration of data. Accessibility was seen as a general quality attribute for public institutions which included the previously mentioned web and physical environment and the addition of tourism.

Group 3 suggested nine additional quality attributes but chose not to link them to specific sectors. They are satisfaction, acceptability, appeal, fun, usefulness, memorability, identification, stimulation and evocation.

Like group 3, group 4 thought that some attributes such as effectiveness, efficiency and pleasure could be associated with all sectors. Five additional attributes were suggested. They are privacy, certainty and confidence, perceived effectiveness and perceived efficiency, and, finally, interaction design solution. This attributes can be divided into professional, quality, knowledge, and state of art-ness.

Viewing Ossiannilsson and Landgren (2012) Group 5 added e-learning to a number of existing quality attributes, i.e. effectiveness (termed productivity), sociability (termed participation) and accessibility. The group also suggested new quality attributes, four of which were said to be attributes of the e-learning sectors. The additional sector specific attributes were flexibility, adaptability, engagement and transparency. The group suggested interactivity as a generic quality attribute concerning system functionality and features.

Group 6 added a number of new quality attributes, privacy, certainty (linked to trust) and safety. Social network sites and web(login) needed to have a privacy quality, construction and health care a certainty quality, and health care, transport and construction a safety quality. Sectors were added to already suggested quality attributes. E-commerce was said to need efficiency and effectiveness and health care effectiveness. The group made several suggestions of merging sectors, i.e. stating that automotive and transport were identical, and so were social network sites and online communities.

From this analysis, we notice that experts feel confident in specifying a number of quality attributes to sectors. However, in other cases they looked at them in a broader way, not specifying attributes to any sectors. Also, that sectors are viewed at different granularity, e.g. transport vs. automotive industry, accessibility for all public institutions and online communities and social networks. The quality attributes themselves could be seen to have different terms, e.g. trust vs. certainty. And experts not only specified attributes for interactivity but also for quality of content, data, results of tasks and usefulness (Jónsdóttir Johannessen & Hornbæk, 2013).

The six quality maps have been consolidated by four groups of six or seven members. An example consolidation is presented in **Fig. 4**.

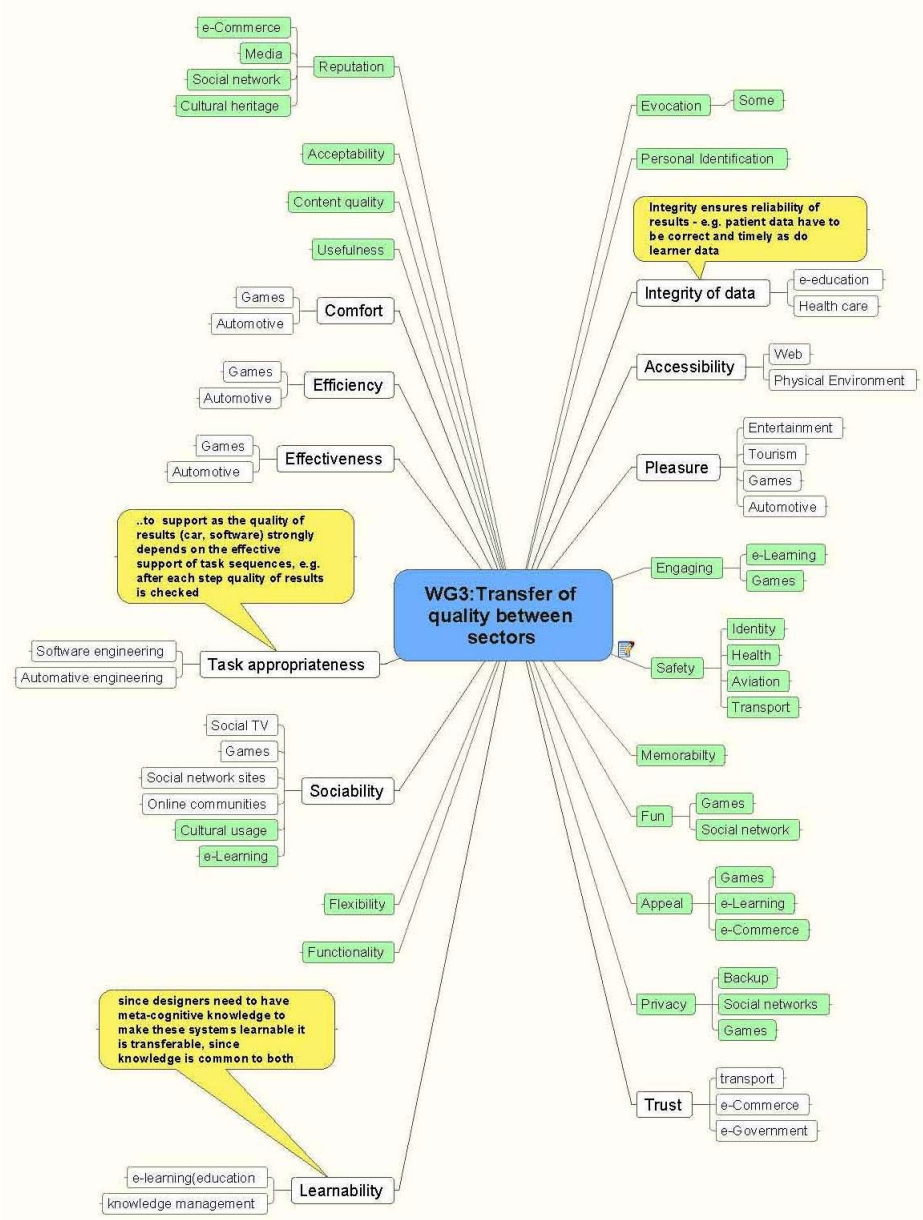



**Fig. 4.** Mapping of quality attributes to sectors[1]

[1] To be translated to an electronic version

**Literature analysis of quality attributes for design and evaluation methods of different sectors.**

The call for contributions on literature analysis of papers that addressed quality attributes while studying design and evaluation methods in a specific sector resulted in 16 papers reviewed. Seven researchers reviewed one or more papers. Out of the 16 papers, two papers did not fully meet the inclusion criteria. Squires and Preece (1999) did not have empirical data and Ossiannilsson and Landgren (2012) mainly discussed quality attributes of methods and did have little empirical data. Nonetheless, they are included in the summary below.

The publications included in this review cover a wide range of sectors/domains. The comparison reveals that while there is some overlap of quality attributes across sectors quality attributes also may also differ within each sector.

Several publications are addressing the e-learning sector. Most of them are using usability inspection methods, such as heuristic evaluation and aim to support the work of evaluators during the inspection, by providing structured guidance in the form of evaluation patterns, specific heuristics, or specific usability guideline.

For instance, Lanzilotti, Ardito, Costabile, and De Angeli (2011) present a pattern-based (PB) usability inspection method, which aims at reducing subjective variability by introducing a set of evaluation patterns to systematically drive inspectors in their evaluation activities. The study showed that evaluation patterns, capitalizing on the reuse of expert evaluators' know-how, provide a systematic framework which reduces reliance on individual skills, increases inter-rater reliability and output standardization, permits the discovery of a larger set of different problems and decreases evaluation cost. The authors explicitly state that the PB inspection can be applied in different contexts on condition that specific evaluation patterns for that domain are defined.

The authors address the e-learning domain by considering both technological and pedagogical aspects. An obvious factor refers to educational quality, where responses such as engagement, motivation, and feeling of control become of fundamental importance. The quality attributes considered are overall quality of e-learning systems, quality in use (e.g. availability of communication tools), and quality of graphical interface elements.

Following a similar idea, Squires and Preece (1999) propose an approach that provides predictive evaluation guidelines for teachers, which can judge the suitability of the software for its educational purposes. The paper presents a set of heuristics, called ''learning with software'', that are an adaptation of the classical usability heuristics, proposed by Jacob Nielsen (1994), for evaluating e-learning systems. This adaptation takes into account the socio-constructivist learning principles, which for many researchers are the basis for developing better educational experiences with the use of software systems. They included specific principles such as match between designer and learner models, navigational fidelity, appropriate levels of learner control, strategies for cognitive error recognition, and match with the curriculum.

The authors stress the importance to conduct predictive evaluations for evaluating the suitability of the educational software. There is a presentation of the relationship between usability heuristics and learning evaluation heuristics (cognitive and

contextual authenticity). Cognitive authenticity consists of credibility, complexity and ownership. Contextual authenticity consists of collaboration and curriculum. Theoretically, the paper is based on socio-constructivist view of learning to identify salient learning issues that should feature in educational software evaluation.

Colla J. MacDonald and Thompson (2005) address the need for quality of e-Learning experiences. They used the Demand-Driven Learning Model (DDLM) (Colla J MacDonald, Stodel, Farres, Breithaupt, & Gabriel, 2001) and its companion evaluation tool (Colla J MacDonald, Breithaupt, Stodel, Farres, & Gabriel, 2002) to evaluate an online Masters in Education course. Multiple data collection methods were used to gather information helping to understand the experiences of stakeholders in this case study: the learners, design team, and facilitators. They found that all five dimensions of the model (structure, content, delivery, service, and outcomes) must work in concert to implement a quality e-Learning course. In addition to adding to the growing knowledge of online learning, their findings highlight additional elements that could be incorporated into the DDLM to further refine the model.

Ginns and Ellis (2007) extend previous research into the domain of blended learning, by exploring the relations between student perceptions of the e-Learning environment, approaches to study, and student grades. The authors used a questionnaire. The aim was to identify a coherent set of scales, with minimal cross-loadings between latent factors. Student responses across the 31 items did not appear to cohere according to the postulated eight scale factor structure (Good e-Teaching, Good e-Resources, Student Interaction and Engagement, Clear Goals and Standards, Appropriate Assessment, Appropriate Workload, Student Management, and Blended Learning). However, using principal axis estimation with oblique rotation to simple structure, by eliminating items with low loadings (less than 0.4) and/or cross-loadings between factors, a subset of eighteen items with a clear factor structure across four factors was identified. These factors were labelled Good e-Teaching, Good e-Resources, Student Interaction, and Appropriate Workload.

Pipan, Arh, and Jerman Blažič (2010) deal with a complex decision-making problem, the selection and evaluation of Learning Management Systems (LMS) in which several objectives - referring to the definite group of users - like social, technical, environmental, and economic impacts, must be simultaneously taken into account. They introduce Evaluation Cycle Management (ECM), a support methodology aimed at the evaluation of options that occur in the decision-making processes. ECM is based on Multi-attribute decision making (Criteria Evaluation) and Usability Evaluation (UX).

The Multi-attribute decision making in the first phase of ECM presents an approach to the development of a qualitative hierarchical decision model that is based on DEX, an expert system shell for multi-attribute decision support. The second phase of ECM is aimed at Usability Evaluation on end users. ECM illustrates its usefulness by showing its main features and its application to the above problem. It is based on the theoretical and practical expertise related to the quality and usability assurance of LMS. The research was focused on three main quality attributes: Student's learning environment (Ease of use, Communication, Functional environment and Help), System, technology & standards category (Technological independence, Security and

privacy, Licensing & hosting and Standards support) and Tutoring & didactics (Course development, Activity tracking and Assessment criteria).

All of the publications concerned with method transfer in the e-Learning sector refer to both generic quality attributes such as ease of use as well as sector specific quality attributes such as learning outcomes or quality of content. Ease of use, communication, collaboration and student interaction are recurring themes across publications.

Another sector of interest for evaluators is entertainment. Two papers targeting video games and digital television were analysed. Both used heuristic evaluation and introduced specific heuristics aimed to support both designer and evaluators.

Pinelle, Wong, and Stach (2008) introduce a set of heuristics based on game reviews that are offering the possibility to evaluate game usability without reviewing unnecessary technical issues and issues related to entertainment, e.g. playability, fun and enjoyment. The paper describes a preliminary evaluation of the heuristics. Five people used them to evaluate a demo version of a PC game.

The authors declare that the goal was to evaluate their overall usefulness at identifying usability problems in games. They also wanted to determine whether knowledgeable evaluators would be able to understand and operationalize the heuristics, and whether the level of detail provided in each was appropriate. There are several similarities between their heuristics and Nielsen's heuristics (Nielsen, 1994). The main difference is that their heuristics use language that makes the links to the game domain more explicit. The paper addresses the game application domain, in particular videogames.

Geerts and Grooff (2009) present twelve sociability heuristics for evaluating social TV, based on several user studies with social TV systems. They performed a competitive analysis of several social television systems as the basis for creating the sociability heuristics, and decided to use an approach based on Grounded Theory for analysing the results. The term 'sociability' is used to indicate these interface aspects that support and enhance social interaction with and through new technologies and applications. The method is based on a heuristic evaluation approach. The quality attributes include free communication, awareness, user control, privacy and minimal distraction.

Relevant quality attributes were also found in papers addressing systems based on different technologies, such as web, virtual reality, and tangible interfaces.

Jokela (2010) studied how to determine usability requirements into a call-for-tenders of a healthcare system. Two quality attributes are addressed: task completion success rate for defining effectiveness requirements, and a complementary measure that we call design solution success rate. A measure for effectiveness (usability) is defined for system requirements: task completion success rate. Effectiveness was seen as a relevant quality attribute of the system. A measure for a 'mixed' quality attribute is defined, covering both efficiency and satisfaction: design solution success rate. This was seen as a relevant quality attribute of the system. The method used is requirements analysis.

Mihajlov and Jerman-Blažič (2011) explore the users' perception of recognition-based, graphical authentication mechanisms in a web environment and aim to find,

whether the memorability of recognition-based authentication keys is influenced by image content. They also examine how the frequency of use affects the usability of the system and whether user training via mnemonic instructions improves the graphical password recognition rate. The sector is security, sub-domain: authentication mechanism. The paper does only in a limited way address a method of evaluation or design. Rather, it discovers properties of designs, e.g. length of authentication, and the effect of knowledge of users, e.g. mnemonic training, characteristics of usage, e.g. frequency of use, and characteristics of users, e.g. gender on quality characteristics such as performance and memorability.

Alistair Sutcliffe and Gault (2004) present a heuristic method for evaluating virtual environment user interfaces. It is based on Nielsen's heuristics (Nielsen, 1994), extended by VE-specific principles proposed by AG Sutcliffe and Kaur (2000). Attributes considered are usability and presence. The domain addressed is virtual environment and applications which were tested were exploring a crime scene and a chess game. The method used is a specialisation, from a generic HE to a specific one for a VE.

Vanden Abeele, Zaman, and De Grooff (2010) suggest Laddering as a promising empirical method to evaluate the impact of tangibility on young children's user experiences. They discuss the typical Laddering interviewing technique and focus on the Laddering data treatment. They argue why Laddering might be especially valuable in a context of UX evaluations of tangible and embedded interfaces for children.

Overall, many of the generic quality attributes mentioned in the publications in this literature review are similar to or even refer back to the original usability heuristics proposed by Nielsen (1994). While some authors adopted them directly, others developed the heuristics further or adapted them to specific sectors or types of systems.

Most publications also mention some quality attributes that are highly specific for a sector such as learning outcomes in e-Learning. Any method transfer will have to consider these sector specific quality attributes. Methods that are open for the integration of such quality attributes by design will be easier to transfer than more rigid methods.

While this study has revealed how quality attributes are applied within the learning sector, further work could use the same protocol to investigate the use of quality attributes in other sectors. Such a follow up study would be able to better compare between sectors whether there are differences in how quality attributes are adopted or adapted. Sectors that might be relevant are the health sector, business and entertainment. The current study has shown that not only do sectors influence the use of quality attributes, but also technologies, such as virtual environment. A follow on study could look at other much investigated technologies such as he much researched mobile technologies (Coursaris & Kim, 2011; Kjeldskov & Paay, 2012) for comparison.

Partly motivated by this work Hvannberg, Halldórsdóttir, and Rudinsky (2012) did a study of how the heuristics of Sutcliffe and Gault for virtual environments had been applied and evolved. 50 papers were analysed and he results showed that a fifth of the papers citing the heuristics have used them fully or partly, and that researchers have

adapted it to their current needs. The analysis of the papers showed that they represent many different application domains, such as manufacturing, psychology treatment and tourism. The way of working, e.g. collaboration, is fairly apparent. Different technologies, such as 3D displays and haptic peripherals could be seen.

**A comparison between quality mapping by experts and systematic literature review.**

There has been no systematic analysis of the quality attribute mapping to sectors by experts and the results of the systematic literature review. One should note that the two activities are not independent since the experts carried out the quality attribute mapping after discussing a subset of the papers from the systematic literature review. Nonetheless, there are some obvious similarities. Both conclude that there are quality attributes specific to sectors. Both conclude that there are quality attributes for various aspects, e.g. interaction or content. In the experts quality mapping implementation technologies are less apparent. The only examples are web and physical environments.

### 3.2 WG3c Use of methods across sectors

There were two activities that aimed to understand how methods could be transferred from one setting to another, especially with respect to quality attributes.

**Story telling of transfer using a lean format .**

At the initial story telling meeting where the form of stories was very short, 20 stories were collected. The form participants were asked to use was: As a <role> I want to <task> so that I can do <need> (see **Fig. 5**). After the stories had been classified and cleaned, a set of 12 stories remained that were related to some kind of transfer of method from one method application to another. Included in the set were stories that were applicable to method transfer and a new form was suggested: Using a method from <source domain> to help me (create/use a method) in the <target domain> to achieve a goal (e.g. ensuring quality). Stories that were excluded were e.g. on gaining, applying or teaching knowledge, selection of a method and method applications.

The analysis revealed several method transfer operations:

- Generalization to specialization, e.g. generic heuristics to specific heuristics for games
- Specialization without any specific reference to a generic method
- Transfer between disjoint domains
- Transfer between similar domains
- Adoption of a technique from a discipline. This adoption may involve change (i.e. adaption) of the method.

The distinction between transfer between two domains and adoption is that the former is from domains, e.g. games, cultural heritage, but adaptation is rather from a

discipline or a body of knowledge. Note, that the distinction may be small in some cases.

To help understand cases of need for transfer of methods, the twelve cases were analysed w.r.s.t. source domain, target domain, a method and quality attributes (**Table 2**).

**Table 2.** 12 stories of transfer classified according to source and target domain, quality attributes and methods that need to be transferred

| **Transfer classification** | **Story** | **Source domain or discipline (if distinct from HCI/ICT)** | **Target domain** | **Quality attribute** | **Method that is for transfer** |
|---|---|---|---|---|---|
| Generalization to specialization | As a designer of an e-learning application I want to use accessibility guidelines to improve the accessibility of my applications | Generic | e-learning | Accessibility | Accessibility guideline |
| | As a web shop tester I want to apply heuristics evaluation from decision making so that I can be sure that my users are supported with their decision making procedures | Decision making | Web shop | Support | Heuristics evaluation |
| Specialization | As a consultant I want to produce a subset of general guidelines so that I can guide the designers to produce interactive solutions of an ERP system | None | ERP systems | None (interactive solutions) | General guidelines |
| Between domains | As a serious game designer I want to use collaboration techniques from e-learning applications to improve the effectiveness of my developed games | e-learning | Games | Effectiveness | Collaboration techniques |
| | As a teacher (e-learning) / designer, I want to explore animations techniques from certain games so that I can increase learner's engagement in exploring content in a reflected way | Games (specific ones) | e-learning | Give the user the possibility to reflect ( Reflectiveness) Engagement | Animations techniques |
| | As a designer of cultural heritage guides I want to give visitors the possibility to play real-time collocated games through their mobile phones so that I can improve the social aspects of the visits | None | Cultural heritage and Games on mobile phones | Sociability | None specified |
| | As an author (cultural heritage) (designer) I want to apply participatory design methods to allow writer-generation discourse for system prototyping | None | Cultural heritage | No quality but functionality (writer-generation discourse) | Participatory design |
| Between similar domains | As a construction engineer I want to know about safety regulations from transport so that I am sure nobody gets hurt when using a very fast elevator | Transport | Elevators | Safety | Safety regulations |

| | | | | | |
|---|---|---|---|---|---|
| Adoption | As a product designer I want to use principles/guidelines for sensory ergonomics so that I can create clear designs for the visually impaired | Sensory field | Visually impaired | Clarity | Guidelines for sensory ergonomics |
| | As a digital TV service designer I want to use lifestyles probes to gain understanding of possible service requirements in domestic environment | None | TV services Domestic environment | None explicit (but could be fit for purpose/ usefulness) | Lifestyle probes |
| | As a usability professional designing interactive and tangible children‘s toys, having found laddering from the marketing domain I need to know what domain aspects method need to be changed to apply the method to evaluate the affective responses to specific attributes of prototypes | Marketing | Toys | Affective response | Laddering to evaluate affective responses |
| | As an e-commerce designer I want to use deep probing of user reactions to prototypes to gain a better understanding of market segments, attitudes and preferences | None | e-commerce | Tailorability to attitudes and preferences | Deep probing of user reactions to prototypes |

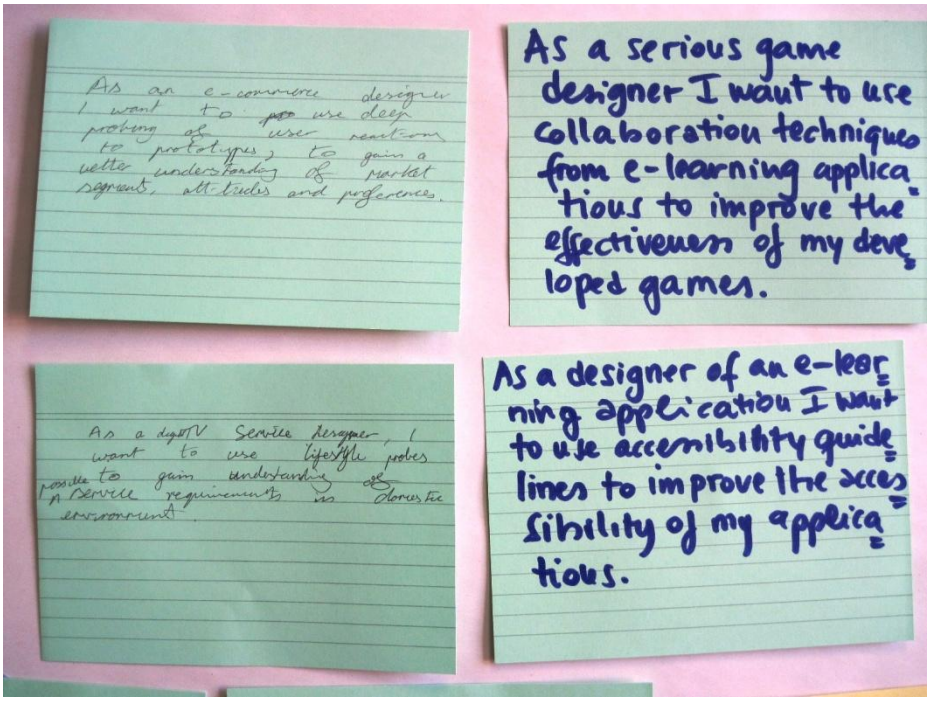



**Fig. 5.** Example stories of transfer

Since eight out of the 20 stories had to be omitted from the analysis, this case study shows that it may be quite difficult for experts to formulate stories of method transfer. Even in the 12 stories, in the quality attribute part there was no specific value of the method transfer and in one case no specific method was mentioned. Setting this aside, i.e. from the remaining 12 stories, the conclusion from this activity is that both the initial format and the revised format can be well used for writing lean stories on method transfer. The transfer classification may need to be improved so that their meaning is clearer. Evaluating the reliability of the classification may also be needed. The results of the analysis shows that the source domain correctly includes domains, e.g. games, transport etc., but also disciplines such as marketing or sensory ergonomics. The targets are clearer, since, understandably, story tellers have a clearer picture of what they are building. In most cases story tellers specified a domain, but there were exceptions. A more technological oriented response was mobile technologies and a user-specific response was visually impaired. Therefore, method and product developers are expected to express the category in which their product falls in various ways.

The quality attributes vary. In one or two cases none could be derived from the story (typically coming from the value/objective part, i.e. so that ….) and in other cases it concerned a quality attribute of the development method (i.e. so that I can *guide the designers* to produce interactive solutions…, or, to allow writer-generation

discourse for system prototyping). In a few cases no quality attribute was made explicit but a functionality of the system. Examples were, so that I can be sure that my users are supported with their *decision making procedures,* or, to gain understanding of possible service requirements.

**Storytelling free formatting and eliciting method resources by experts.**

Motivated by the success of storytelling of transfer in the London workshop, TwinTide members were asked to participate in story telling but in a slightly different manner. Instead of restricting them to a specific format they were asked to tell two stories of method applications, e.g. in two separate projects on an index card. One story on each side. These stories were written and then refined in a subsequent meeting. After the initial exercise 18 story tellers submitted one or two stories and classified their transfer. The stories were transcribed to a document and the story tellers were asked to correct them. Two researchers reviewed the 18 cases and concluded that 4 of them were not fit for further analysis because of lack of information. Of the 14 remaining cases three only had one story but we were interested in cases where at least pairs were presented. Therefore, 11 cases were analysed.

This section shows three example cases but other cases are listed in Appendix B: "Stories from Experts". Case 2 which has examples from a 3D application and an AR-application are provided in **Table 3** with an illustration of how different method resources are used in **Table 4**.

**Table 3.** Case 2 Methods adapted to a 3D application and to an AR-application

| **Case 2** | | | |
|---|---|---|---|
| **Subcase** | **Domain** | **Method** | **Transfer** |
| 1 | 3D, Simulation | Mix user testing with expert evaluation guided by ergonomic criteria for 3D application | Adaptation |
| 2 | AR-application for museum. Classification of "life" | Two hour Expert session with usability engineers, archivation experts, HCI experts, users, museum experts. Working on models | Adaptation |

The storytellers noted that Analysis as a method resource is more focused. Explanation or hints for redesign could require additional effort. Transfer notes on participant recruitment were explained, saying users needed to be recruited to identify problems/solutions, but might not contribute to redesign

**Table 4.** Method resources contrasted for two subcases of Case 2

| **Case 2** | | |
|---|---|---|
| **Method resources** | **Subcase 1** | **Subcase 2** |
| Task selection | | Focused tasks given by application |
| Problem identification | Made by user testing suggested by ergonomic criteria | Rationale identification |
| Analysis | Users required… background<br>Ergonomic criteria – role guiding process | Task specific<br>Rationale of features could be identified |
| Participant recruitment | User recruitment for user testing. Users could not identify actual problems. No background knowledge -> gave precedent to prediction | User sessions in the course of co-evaluation. Results in terms of satisfying users. No explanation why it worked |
| Method reflection/experience | UI description is important<br>Categorization is efficient<br>User testing efficient to find problems | |
| Problem classification | Ergonomic criteria for classifying problems<br>Single | Elicitation of expert knowledge for classification of life<br>Many people |
| Problem identification | Inspection: description, user testing<br>Single | Co-evaluation consisting of an expert session<br>Many people |

The next case we show is a generalization to specialization of a questionnaire. The case is shown in **Table 5**. No method resources were specified.

**Table 5.** Case 5 Generalization to specialization

| **Case 5** | | | |
|---|---|---|---|
| **Subcase** | **Domain** | **Method** | **Transfer** |
| 1 | Math problem solving | Questionnaire on how competent you feel on the problem solving task on a scale of 1-10<br>Answer on paper | |
| 2 | Educational game | You cannot interrupt the flow of game. A non-player character was created who presented the same format question to the player in | Generalization to specialization, Adaptation applies also. |

| | | | |
|---|---|---|---|
| | | text format. The player replied by clicking a number.<br>User tasks are different in this context from 1. | |

The final example we present is from case 10 that includes using task models as a method. The case is presented in **Table 6** and method resources in **Table 7**.

Subcase 1 of case 10 has been reported in (Winckler, Bach, & Bernhaupt, 2012). Subcase 2 has been published in (ter Beek, Faconti, Massink, Palanque, & Winckler, 2009) and (Palanque et al., 2009).

**Table 6.** Case 10 Task modeling

| **Case 10** | | | | |
|---|---|---|---|---|
| **Sub-case** | **Domain** | **Method** | **Transfer** | **Quality attribute** |
| 1 | Mobile application for citizens to report incidents | 1. First interviews (structured) current situation). Grounded theory method for analysis UX dimensions in the stories (& context). Memories & wishes. Based on this, the model was generated.<br>2. Task analysis (by researchers) From the wishes collected in step 1, analyse the steps that participants would like to do to accomplish a task.<br>3. Field test (predefined context on campus) - This is where model based task analysis was used.<br>The task model developed in the earlier phases was used to see if the UX dimensions in the tasks matched the original wishes. | | |
| 2 | Office work / Organization of icons on desktop. | There was a given model of the task. Controlled experiment about user performance study in a lab, see if the model holds true.<br>Ask participants to perform simple task on desktop (dragging icon to trash), calculate times of each part of the task as described in the model (with and without interruption), | | User performance in time |

**Table 7.** Method resources of case 10

| **Case 10** | | |
|---|---|---|
| **Method resource** | **Subcase 1** | **Subcase 2** |
| Modeling | Concurrent Task Trees (CTT)(Paternò, Mancini, & Meniconi, 1997) | Concurrent Task Trees (CTT) |
| Task analysis | Iterative process | Iterative process |
| Task modeling process | Iterative process | Iterative process |
| Participant recruiting | N/A researchers were doing the model-based task analysis, the same researchers in both cases | N/A researchers were doing the model-based task analysis, the same researchers in both cases |
| Task selection | The tasks modelled in the model were different | The tasks modelled in the model were different |
| Problem identification | done against the model | done against the model |
| Problem classification | done against the model | done against the model |
| Reporting format | In UX test, qualitative results. The task model itself was used in both cases to visualize the results. | In performance test, quantitative findings. |
| Analysis | Done against the model | Done against the model |

Five of the cases did contrast method resources in the two subcases. Seven cases reported on a classification of transfer. In three of the cases, this researcher could extract quality attributes.

To learn more about the quality attributes aimed for in the case studies it was planned to inquire storytellers about their use in the case studies. The survey template is in Appendix C: “A Survey on Quality Attributes”.

# 4 Impacts Achieved and or projected

## 4.1 List of publications

The main contribution of WG3 to publications were to the workshop held at CHI'13 Made for sharing, HCI stories on transfer, tragedy and triumph (Law et al., 2013) and a specific contribution therein (Hvannberg, 2013) on transfer along the sensory and collaborative axes. Finally, the work of WG3 influenced a paper by Hvannberg on identifying and explicating knowledge on method transfer using a sectorial system of innovation approach to be published in a special issue of UAIS (see the final report by WG2, this report).

## 4.2 List of outreach activities

As mentioned above an open workshop was held at CHI'13.

## 4.3 Achievements and projections

The main contribution of the working group has been to encourage experts to elicit empirical experiences on the use and adaption of methods for design and evaluation, in particular taking into account quality attributes. Such descriptions and hence reflection can be a major motivation for advances in the field of transfer of methods. The frameworks, using quality attribute maps, the stories, method resources contrasting and the literature analysis can be tools for an analysis of these experiences and reflections. With the explication of knowledge we have encouraged dialogues between researchers in an informal way, giving ideas on how to study the literature and encourage practitioners of HCI to report on their experiences. Researchers can also serve as a medium for a conversation with practitioners, helping them to evoke their experiences.

# 5 Conclusion and future work

The TwinTide action has been an excellent venue for researchers to contribute to description and reflection on methods applied in different contexts and sectors. Overall, many of the generic quality attributes mentioned in the publications in the literature review are similar to or even refer back to the original usability heuristics proposed by Nielsen (1994). While some authors adopted them directly, others developed the heuristics further or adapted them to specific sectors or types of systems. Similar results were derived by (Hvannberg et al., 2012).

Most publications also mention some quality attributes that are highly specific for a sector such as learning outcomes in e-Learning. Any method transfer will have to consider these sector specific quality attributes. Methods that are open for the integration of such quality attributes by design will be easier to transfer than more rigid methods.

It should be emphasized that with this work we are only just beginning to explicate and reflect on the transfer and adaption of methods. We have learned what experts can provide and what they may have difficulties in answering. Future work should emphasize developing more such tools to help them reflect on their use of methods and how they develop them for new sectors.

# 6 References


1. Coursaris, C. K., & Kim, D. J. (2011). A meta-analytical review of empirical mobile usability studies. *Journal of usability studies, 6*(3), 117-171.
2. DIS, I. (2010). 9241-210: 2009. Ergonomics of human system interaction-Part 210: Human-centred design for interactive systems (formerly known as 13407). *International Organization for Standardization (ISO). Switzerland*.
3. Geerts, D., & Grooff, D. D. (2009). *Supporting the social uses of television: sociability heuristics for social tv*. Paper presented at the Proceedings of the 27th international conference on Human factors in computing systems, Boston, MA, USA.
4. Ginns, P., & Ellis, R. (2007). Quality in blended learning: Exploring the relationships between on-line and face-to-face teaching and learning. *The Internet and Higher Education, 10*(1), 53-64. doi: 10.1016/j.iheduc.2006.10.003
5. Hvannberg, E. T. (2013). *Impact of Sensory and Collaborative Contexts on Usability Evaluation Methods Resources*. Paper presented at the CHI'13 Workshop. Made for sharing: HCI stories of transfer, triumph and tragedy
6. Hvannberg, E. T., Halldórsdóttir, G., & Rudinsky, J. (2012). *Exploitation of heuristics for virtual environments.* Paper presented at the Proceedings of the 7th Nordic Conference on Human-Computer Interaction: Making Sense Through Design.
7. Jokela, T. (2010). *Determining usability requirements into a call-for-tenders: a case study on the development of a healthcare system*. Paper presented at the Proceedings of the 6th Nordic Conference on Human-Computer Interaction: Extending Boundaries, Reykjavik, Iceland.
8. Jónsdóttir Johannessen, G. H., & Hornbæk, K. (2013). Must evaluation methods be about usability? Devising and assessing the utility inspection method. *Behaviour & Information Technology*, 1-12. doi: 10.1080/0144929x.2012.751708
9. Kitchenham, B. A., & Charters, S. (2007). Guidelines for performing systematic literature reviews in software engineering.
10. Kjeldskov, J., & Paay, J. (2012). *A longitudinal review of Mobile HCI research Methods.* Paper presented at the Proceedings of the 14th international conference on Human-computer interaction with mobile devices and services.
11. Lanzilotti, R., Ardito, C., Costabile, M. F., & De Angeli, A. (2011). Do patterns help novice evaluators? A comparative study. *International Journal of Human-Computer Studies, 69*(1–2), 52-69. doi: 10.1016/j.ijhcs.2010.07.005
12. Law, E. L.-C., Hvannberg, E. T., & Cockton, G. (2008). A green paper on usability maturation *Maturing Usability* (pp. 381-424): Springer.

13. Law, E. L.-C., Hvannberg, E. T., Vermeeren, A. P., Cockton, G., & Jokela, T. (2013). *Made for sharing: HCI stories of transfer, triumph and tragedy.* Paper presented at the CHI'13 Extended Abstracts on Human Factors in Computing Systems.
14. MacDonald, C. J., Breithaupt, K., Stodel, E. J., Farres, L. G., & Gabriel, M. A. (2002). Evaluation of web-based educational programs via the Demand-Driven Learning Model: A measure of web-based learning. *International Journal of Testing, 2*(1), 35-61.
15. MacDonald, C. J., Stodel, E. J., Farres, L. G., Breithaupt, K., & Gabriel, M. A. (2001). The demand-driven learning model: A framework for web-based learning. *The Internet and Higher Education, 4*(1), 9-30.
16. MacDonald, C. J., & Thompson, T. L. (2005). Structure, Content, Delivery, Service, and Outcomes: Quality e-Learning in higher education. *The International Review of Research in Open and Distance Learning, North America, 6*(2), 1-14.
17. Mihajlov, M., & Jerman-Blažič, B. (2011). On designing usable and secure recognition-based graphical authentication mechanisms. *Interacting with Computers, 23*(6), 582-593. doi: 10.1016/j.intcom.2011.09.001
18. Nielsen, J. (1994). Heuristic evaluation. In N. Jakob & L. M. Robert (Eds.), *Usability inspection methods* (pp. 413): John Wiley & Sons, Inc.
19. Ossiannilsson, E., & Landgren, L. (2012). Quality in e - learning - a conceptual framework based on experiences from three international benchmarking projects. *Journal of Computer Assisted Learning, 28*(1), 42-51.
20. Palanque, P., Winckler, M., Ladry, J.-F., ter Beek, M. H., Faconti, G., & Massink, M. (2009). *A formal approach supporting the comparative predictive assessment of the interruption-tolerance of interactive systems.* Paper presented at the Proceedings of the 1st ACM SIGCHI symposium on Engineering interactive computing systems.
21. Paternò, F., Mancini, C., & Meniconi, S. (1997). *ConcurTaskTrees: A diagrammatic notation for specifying task models.* Paper presented at the Human-Computer Interaction INTERACT'97.
22. Pinelle, D., Wong, N., & Stach, T. (2008, April 5-10, 2008). *Heuristic Evaluation for Games:Usability Principles for Video Game Design.* Paper presented at the CHI '08, CHI Conference on Human Factors in Computing Systems Florence, Italy.
23. Pipan, M., Arh, T., & Jerman Blažič, B. (2010). The Evaluation Cycle Management Method Applied to the Evaluation of Learning Management System *Integrating usability engineering for designing the web experience: methodologies and principles.*: Hershey Information Science (IGI-Global).
24. Raskin, J. (2000). *The humane interface: new directions for designing interactive systems*: Addison-Wesley Professional.
25. Squires, D., & Preece, J. (1999). Predicting quality in educational software:: Evaluating for learning, usability and the synergy between them. *Interacting with Computers, 11*(5), 467-483. doi: 10.1016/s0953-5438(98)00063-0
26. Sutcliffe, A., & Gault, B. (2004). Heuristic evaluation of virtual reality applications. *Interacting with Computers, 16*(4), 831-849.

27. Sutcliffe, A., & Kaur, K. D. (2000). Evaluating the usability of virtual reality user interfaces. *Behaviour & Information Technology, 19*(6), 415-426.
28. ter Beek, M. H., Faconti, G. P., Massink, M., Palanque, P. A., & Winckler, M. (2009). Resilience of interaction techniques to interrupts: A formal model-based approach *Human-Computer Interaction–INTERACT 2009* (pp. 494-509): Springer.
29. Vanden Abeele, V., Zaman, B., & De Grooff, D. (2010). User eXperience Laddering with preschoolers: unveiling attributes and benefits of cuddly toy interfaces. *Personal and Ubiquitous Computing*, 1-15. doi: 10.1007/s00779-011-0408-y
30. Winckler, M., Bach, C., & Bernhaupt, R. (2012). *Identifying User eXperiencing Factors along the Development Process: A Case Study*. Paper presented at the I-UxSED.

## Appendix A: Complete set of references from London workshop


1. Geerts, D., & Grooff, D. D. (2009). Supporting the social uses of television: sociability heuristics for social TV. Paper presented at the Proceedings of the 27th international conference on Human factors in computing systems, Boston, MA, USA.
2. Ghiani, G., Paternò, F., Santoro, C., & Spano, L. D. (2009). UbiCicero: A location-aware, multi-device museum guide. Interacting with Computers, 21(4), 288-303.
3. Ginns, P., & Ellis, R. (2007). Quality in blended learning: Exploring the relationships between on-line and face-to-face teaching and learning. The Internet and Higher Education, 10(1), 53-64. doi: 10.1016/j.iheduc.2006.10.003
4. Grossman, T., Fitzmaurice, G., & Attar, R. (2009). A survey of software learnability: metrics, methodologies and guidelines. Paper presented at the Proceedings of the 27th international conference on Human factors in computing systems, Boston, MA, USA.
5. Jokela, T. (2010). Determining usability requirements into a call-for-tenders: a case study on the development of a healthcare system. Paper presented at the Proceedings of the 6th Nordic Conference on Human-Computer Interaction: Extending Boundaries, Reykjavik, Iceland.
6. Lanzilotti, R., Ardito, C., Costabile, M. F., & De Angeli, A. (2011). Do patterns help novice evaluators? A comparative study. International Journal of Human-Computer Studies, 69(1–2), 52-69. doi: 10.1016/j.ijhcs.2010.07.005
7. MacDonald, C. J., & Thompson, T. L. (2005). Structure, Content, Delivery, Service, and Outcomes: Quality e-Learning in higher education. The International Review of Research in Open and Distance Learning, North America, 6(2), 1-14.
8. Mihajlov, M., & Jerman-Blažič, B. (2011). On designing usable and secure recognition-based graphical authentication mechanisms. Interacting with Computers, 23(6), 582-593. doi: 10.1016/j.intcom.2011.09.001
9. Mihajlov, M., Jerman-Blažič, B., & Tevdovski, D. (2007). Prototyping mobile graphical authentication. International Review on Computers and Software (I.RE.CO.S.) xx(x), 1-8.
10. Ossiannilsson, E., & Landgren, L. (2012). Quality in e - learning - a conceptual framework based on experiences from three international benchmarking projects. Journal of Computer Assisted Learning, 28(1), 42-51.
11. Pinelle, D., Wong, N., & Stach, T. (2008, April 5-10, 2008). Heuristic Evaluation for Games:
12. Usability Principles for Video Game Design. Paper presented at the CHI '08, CHI Conference on Human Factors in Computing Systems Florence, Italy.
13. Pipan, M., Arh, T., & Jerman Blažič, B. (2010). The Evaluation Cycle Management Method Applied to the Evaluation of Learning Management System Integrating usability engineering for designing the web experience: methodologies and principles.: Hershey Information Science (IGI-Global).
14. Rice, M., Newell, A., & Morgan, M. (2007). Forum Theatre as a requirements gathering methodology in the design of a home telecommunication system for older adults. Behaviour & Information Technology, 26(4), 323-331. doi: 10.1080/01449290601177045

15. Squires, D., & Preece, J. (1999). Predicting quality in educational software:: Evaluating for learning, usability and the synergy between them. Interacting with Computers, 11(5), 467-483. doi: 10.1016/s0953-5438(98)00063-0
16. Sutcliffe, A., & Gault, B. (2004). Heuristic evaluation of virtual reality applications. Interacting with Computers, 16(4), 831-849.
17. Vanden Abeele, V., Zaman, B., & De Grooff, D. (2010). User eXperience Laddering with preschoolers: unveiling attributes and benefits of cuddly toy interfaces. Personal and Ubiquitous Computing, 1-15. doi: 10.1007/s00779-011-0408-y

## Appendix B: Stories from Experts

| **Case 1** | | | |
|---|---|---|---|
| **Subcase** | **Domain** | **Method** | **Transfer** |
| 1 | Health care | Expert evaluation<br>Reporting of problems<br>Proposed solutions | Adaptation |
| 2 | Document and administration management | Expert evaluation<br>15 criteria based on 9241-110 dialogue variables | A new method created to meet client needs |

| **Case 2** | | | |
|---|---|---|---|
| **Subcase** | **Domain** | **Method** | **Transfer** |
| 1 | 3D, Simulation | Mix user testing with expert evaluation guided by ergonomic criteria for 3D application | Adaptation |
| 2 | AR-application for museum. Classification of "life" | Expert session with usability engineers, archivation experts, HCI experts, users, museum experts. Working on models | Adaptation |

| **Case 2** | | |
|---|---|---|
| **Method resources** | **Subcase 1** | **Subcase 2** |
| Task selection | | Focused tasks given by application |
| Problem identification | Made by user testing<br>suggested by ergonomic criteria | Rationale identification |
| Analysis | Users required… background<br>Ergonomic criteria – role guiding process | Task specific<br>Rationale of features could be identified |
| Method reflection/experience | UI description is important<br>Categorization is efficient<br>User testing efficient to find problems | |
| Problem classification | Ergonomic criteria for classifying problems<br>Single | Elicitation of expert knowledge for classification of life<br>Many people |
| Problem identification | Inspection: description, user testing<br>Single | Co-evaluation consisting of an expert session<br>Many people |

| Case 3 | | | | |
|---|---|---|---|---|
| **Subcase** | **Domain** | **Method** | **Transfer** | **Quality attribute** |
| 1 | TV | Questionnaire applied every seven minutes on social connectedness and MOS scores<br>Remote interaction / controlled environment | None | Desynchronization of videos |
| 2 | TV | Explorative study / less controlled<br>Group work, individual work and group interviews | Adaptation | Comparison of types of devices to share video content |

| Case 3 | | |
|---|---|---|
| **Method resource** | **Subcase 1** | **Subcase 2** |
| Problem identification | Normative | Problems / opportunities<br>Open ended / exploratory |
| Reporting format | Qualitative<br>inferential | Mixed quantitative/qualitative<br>Descriptive statements |
| Task selection | Specific task<br>One task | Open ended<br>more tasks |
| Participant recruitment | Recruitment office<br>Requirements: Age, Experience w. computers, Watching TV, All people should know each other | Same as 1 |

| Case 4 | | | |
|---|---|---|---|
| **Subcase** | **Domain** | **Method** | **Transfer** |
| 1 | ePortfolio | deLone and McLean's IS success model | |
| 2 | ePortfolio | Specification of relationships between constructs | Generalization to specialization and adaptation |

| **Case 4** | | |
|---|---|---|
| **Method resource** | **Subcase 1** | **Subcase 2** |
| Personnel recruitment | Experts by individual e-mails | End-users invited by e-mail sent to their organization – mailing list |
| Problem identification | Literature research | |
| Analysis | Excel spread-sheet disseminated by e-mail | On-line questionnaire, end users invited by e-mail |

| **Case 5** | | | |
|---|---|---|---|
| **Subcase** | **Domain** | **Method** | **Transfer** |
| 1 | Math problem solving | Questionnaire on how competent you feel on the problem solving task on a scale of 1-10<br>Answer on paper | |
| 2 | Educational game | You cannot interrupt the flow of game. A non-player character was created who presented the same format question to the player in text format. The player replied by clicking a number.<br>User tasks are different in this context from 1. | Generalization to specialization, Adaptation applies also. |

| **Case 6** | | | | |
|---|---|---|---|---|
| **Subcase** | **Domain** | **Method** | **Transfer** | **Quality attribute** |
| 1 | Website | Rating of satisfaction, time to accomplish tasks, errors, efforts of visited links,<br>Before and after (expected and perceived satisfaction) | | Usability |
| 2 | Mobile tour guide | Usability assessment. Not a desktop computer, not a website but a mobile application.<br>Challenging to evaluation effectiveness, efficiency and satisfaction.<br>Tasks no longer restricted to the application but to the outside world. | Extension | Usability |

| **Case 7** | | | |
|---|---|---|---|
| **Subcase** | **Domain** | **Method** | **Transfer** |
| 1 | Grant practices Standards' assessment in a NGO (ISO9000:1) | Visits to location, interviews, document inspection. | Adaptation |
| 2 | Compliance to law in an Information system | Check list inspection. Interviews. Observations, document inspection was not possible due to lack of documentation. | |

| **Case 8** | | | |
|---|---|---|---|
| **Subcase** | **Domain** | **Method** | **Transfer** |
| 1 | Collaborative synchronous application for Learning | Knowledge related to Groupware evaluation heuristics (mechanics of collaboration). Learning had to be adapted | Adaptation |
| 2 | Mobile game | Created a code for observation of game playing along dimensions which were taken from Game theory and embodied interaction. Social dynamics. Created an evaluation scheme and analysed the video. | |

| **Case 9** | | | |
|---|---|---|---|
| **Subcase** | **Domain** | **Method** | **Transfer** |
| 1 | Home entertainment: Set top box for elderly | Fixed task user testing, with instant data analysis. Daily email reports to developer. Wrote user manual. Out of box testing (digital out of the box testing). 3 week total. Overnight fixes / daily iterations / end of week – until no new problems revealed. Task set charged (2 week) – new problems appeared e.g. account configuration (personal details) connection (email) User groups: single, | |

| | | | |
|---|---|---|---|
| | | pairs, trios were testing groups. Setting: 3 test setting | |
| 2 | Website | Participant-recruitment was outsourced recruitment. Location arrangements by agency (library). Socially excluded people. Problem identification. Fixed task user testing. Technology: laptop and webcam and video recording. Reporting Format. Deliverable was report highlight. | |

Differences between subcases 1 and 2 were regarding relationship to clients – social. In case 1 it was direct but in case 2 it was indirect. Relationship to developers was direct in case 1 and in case 2 there was no contact. Case one was stage setting (lab) but case 2 was goal setting (library).

| **Case 9** | | |
|---|---|---|
| **Method resource** | **Subcase 1** | **Subcase 2** |
| Participant recruitment | direct relationship with the participants through direct recruitment | indirect relationship with the participants through recruitment organised through agent |
| Relationship to developers | direct contact to the developers to ensure an iterative process, flexible setting | no contact at all to the developers, fixed setting |
| Study setting | lab based setting | library / field based setting |

Subcase 1 of case 10 has been reported in (Winckler et al., 2012). Subcase 2 has been published in (ter Beek et al., 2009) and (Palanque et al., 2009).

| **Case 10** | | | | |
|---|---|---|---|---|
| **Sub-case** | **Domain** | **Method** | **Tran sfer** | **Quality attribute** |
| 1 | Mobile application for citizens to report incidents | 1. First interviews (structured) current situation). Grounded theory method for analysis UX dimensions in the stories (& context). Memories & wishes. Based on this, the model was generated. | | |

| | | | | |
|---|---|---|---|---|
| | | 2. Task analysis (by researchers) From the wishes collected in step 1, analyse the steps that participants would like to do to accomplish a task.<br>3. Field test (predefined context on campus) - This is where model based task analysis was used.<br>The task model developed in the earlier phases was used to see if the UX dimensions in the tasks matched the original wishes. | | |
| 2 | Office work / Organization of icons on desktop. | There was a given model of the task. Controlled experiment about user performance study in a lab, see if the model holds true.<br>Ask participants to perform simple task on desktop (dragging icon to trash), calculate times of each part of the task as described in the model (with and without interruption), | | User performance in time |

| **Case 10** | | |
|---|---|---|
| **Method resource** | **Subcase 1** | **Subcase 2** |
| Modeling | Concurrent Task Trees (CTT)(Paternò et al., 1997) | Concurrent Task Trees (CTT) |
| Task analysis | Iterative process | Iterative process |
| Task modeling process | Iterative process | Iterative process |
| Participant recruiting | N/A researchers were doing the model-based task analysis, the same researchers in both cases | N/A researchers were doing the model-based task analysis, the same researchers in both cases |
| Task selection | The tasks modelled in the model were different | The tasks modelled in the model were different |
| Problem identification | done against the model | done against the model |
| Problem classification | done against the model | done against the model |
| Reporting format | In UX test, qualitative results. The task model itself was used in both cases to visualize the results. | In performance test, quantitative findings. |
| Analysis | Done against the model | Done against the model |

| **Case 11** | | | | |
|---|---|---|---|---|
| **Sub-case** | **Domain** | **Method** | **Transfer** | **Quality attbute** |
| 1 | Automotive industry | Combination of mobile electronic diary to capture positive and negative experiences with text, video, and audio stories told by the user (self-reporting) | Generalization to specialization | Com-fort. |
| 2 | Building | Diary study combined with post-hoc semi-structured interview that used reports about experiences to recapitulate and discuss | Generalization to specialization | Access-ibility |

Mobile devices were used as for diaries that captured positive negative experiences with auto, video, photo as they were told by the user. The diaries were a basis for semi-structured interviews. Incidents of comfort/discomfort were discussed. Same method or approach ("diary smartphone") was used on a general level, but specialized for different domains by changing core questions.

The main difference between the subcases one and two in case 11 is that the first study is more narrative, gives more subjective experiences and emotions and there are suggestions from users for improvements. In the second study there was more documenting, descriptive, more use of pictures but less narrative.

## Appendix C: A survey on Quality Attributes

For each of the case studies you have reported, please indicate which quality in use characteristics were used as goals of the usability evaluation

| | |
|---|---|
| **Effectiveness** | |
| **Efficiency** | |
| **Satisfaction**: | |
| | Usefulness |
| | Pleasure |
| | Comfort |
| **Freedom from risk:** | |
| | Economic risk mitigation |
| | Health and safety risk mitigation |
| | Environmental risk mitigation |
| **Context coverage:** | |
| | Context completeness |
| | Flexibility |
| **Other quality characteristics (name one or more)** | |

**1.1 effectiveness**
accuracy and completeness with which users achieve specified goals
**1.2 efficiency**
resources expended in relation to the accuracy and completeness with which users achieve goals
**1.3 satisfaction**
degree to which user needs are satisfied when a product or system is used in a specified context of use
**1.3.1 usefulness**
degree to which a user is satisfied with their perceived achievement of pragmatic goals, including the results of use and the consequences of use
**1.3.2 trust**
degree to which a user or other stakeholder has confidence that a product or system will behave as intended
**1.3.3 pleasure**
degree to which a user obtains pleasure from fulfilling their personal needs
**1.3.4 comfort**
degree to which the user is satisfied with physical comfort
**1.4 freedom from risk**

degree to which a product or system mitigates the potential risk to economic status, human life, health, or the environment

NOTE Risk is a function of the probability of occurrence of a given threat and the potential adverse consequences of that threat's occurrence.

**1.4.1 economic risk mitigation**

degree to which a product or system mitigates the potential risk to financial status, efficient operation, commercial property, reputation or other resources in the intended contexts of use

**1.4.2 health and safety risk mitigation**

degree to which a product or system mitigates the potential risk to people in the intended contexts of use

**1.4.3 environmental risk mitigation**

degree to which a product or system mitigates the potential risk to property or the environment in the intended contexts of use

**1.5 context coverage**

degree to which a product or system can be used with effectiveness, efficiency, freedom from risk and satisfaction in both specified contexts of use and in contexts beyond those initially explicitly identified

**1.5.1 context completeness**

degree to which a product or system can be used with effectiveness, efficiency, freedom from risk and satisfaction in all the specified contexts of use

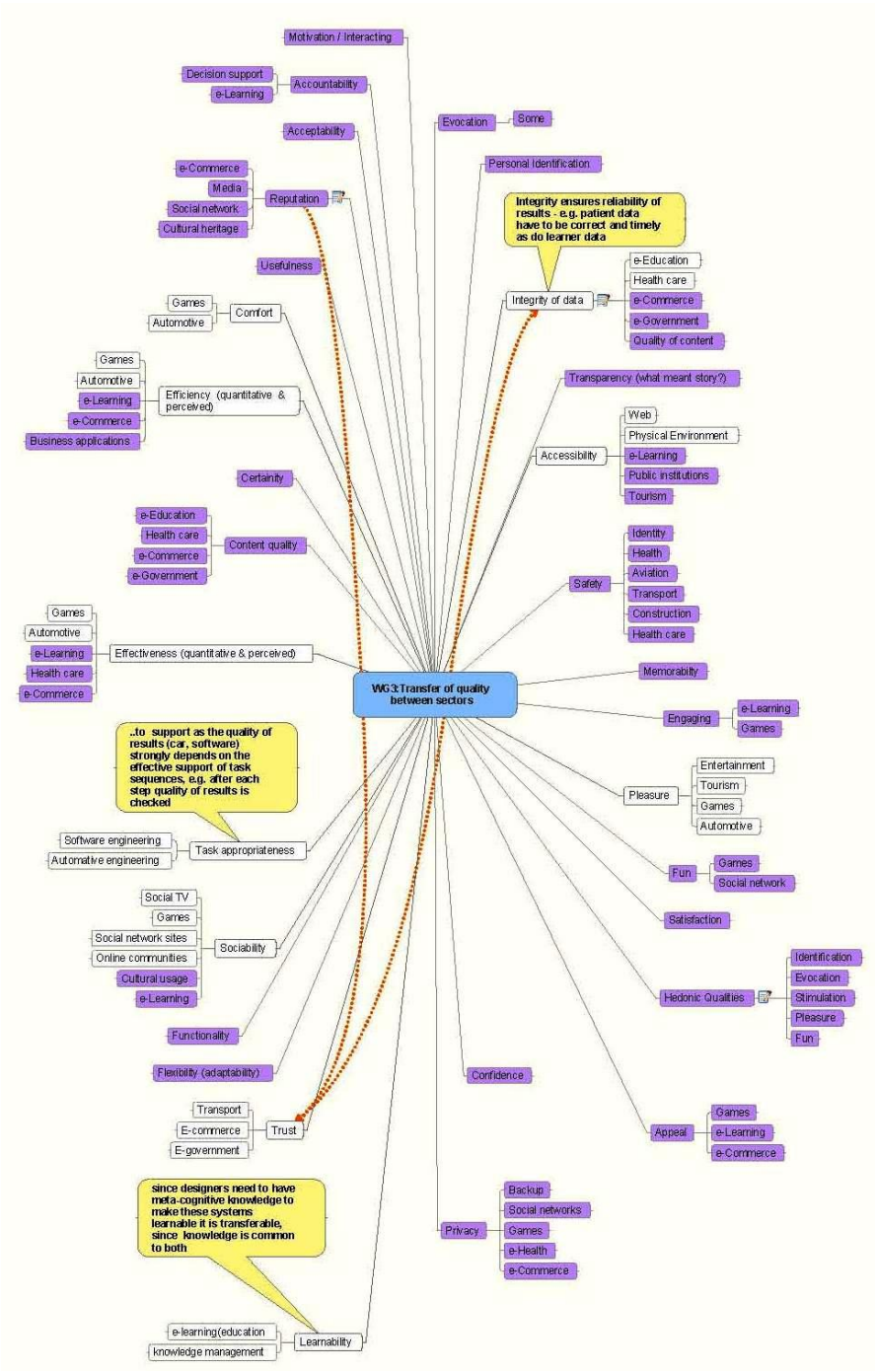
EXAMPLE The extent to which software is usable using a small screen, with low network bandwidth, by a non-expert user; and in a fault-tolerant mode (e.g. no network connectivity).

**1.5.2 flexibility**

degree to which a product or system can be used with effectiveness, efficiency, freedom from risk and satisfaction in contexts beyond those initially specified in the requirements NOTE 1 Flexibility can be achieved by adapting a product (4.2.8.1) for additional user groups, tasks and cultures.

NOTE 2 Flexibility enables products to take account of circumstances, opportunities and individual preferences that may not have been anticipated in

# Appendix D: Consolidation of Quality Attributes per Sector



# TwinTide WG4 Report

## Framework for Diffusion of Design & Evaluation Method Innovation (DIDEMI)

**Arnold P.O.S Vermeeren**

TU Delft, the Netherlands

# Framework for Diffusion of Design & Evaluation Method Innovation (DIDEMI) (WG4)

Arnold P.O.S. Vermeeren

Delft University of Technology, Industrial Design Engineering, Landbergstraat 15, NL 2628 CE Delft, The Netherlands.

a.p.o.s.vermeeren@tudelft.nl[1]

**Abstract.** TwinTide WG4 proposes a framework for Diffusion of Design & Evaluation Method Innovations by integrating the findings from WG1, WG2 and WG3. The integrated framework indicates where Multi-Criteria Decision Making Mechanisms would fit in the framework. It also provides a process overview of selecting new methods in an organization and infers which types of factors may influence the process. Essentially the framework is based on Rogers' well-known Diffusion of Innovation approach, but his agenda-setting stage is more detailed, and influencing factors identified in the framework are more tailored to the situation of design and evaluation methods.

**Keywords:** Design, Evaluation, Methods, Diffusion of Innovations, Multi-criteria Decision Making.

## 1 Objectives and Tasks

The aim of WG4 is to develop an integrated Design &Evaluation methods selection framework (hereafter 'framework') that should help practitioners in selecting methods for use in their specific situations, and serve as a reference model for future development of HCI in terms of training, education and identifying opportunities for developing new methods.

This WG's activities consist of two main tasks:

(i) synthesizing a framework of D&E methods based on outcomes of the other three WGs;

(ii) developing decision support mechanisms for selecting D&E methods.

[1] Authors of the section 2.2.3 on MCDM written by Paulo Melo and Bojan Sdrjevic

# 2 Approach

## 2.1 Process summary

WG4 started in 2011, half way the TwinTide project. Until 2011 WG1, WG2 and WG3 had focused on the concepts of method, sector and transfer. Concepts relevant in cases of transfer of methods across sectors were identified. However, for gaining an understanding of and facilitating method selection in practice, a more holistic view is required: not only focusing on theoretical notions of methods, sectors and transfer, but also taking into account multiple other pragmatic factors that play a role in selecting new methods in real-world practitioners' contexts. To that end, WG4 took off with imagining what would be required from a tool that would facilitate method selection: who would use it, what would they use it for, and how would they like to use such a tool. From then on, two tracks were followed around the two tasks of WG4: 1) developing a general framework for method diffusion across contexts, and 2) exploring Multi-Criteria Decision Making Mechanisms to be used along with the framework. See Figure 1 for an overview of the process. In the next sections the two tracks are summarized.

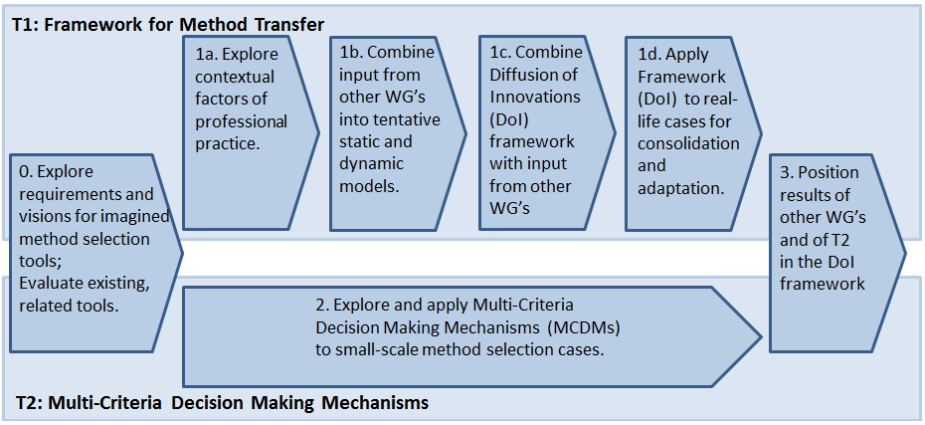



**Fig. 1.** Diagram showing WG4's process model.

### 2.1.1 Framework for method diffusion across contexts

For studying transfer of methods across contexts one can take various perspectives. One perspective to take is to focus on the concerns of individuals wanting to adopt new methods, another would be to take a bird's eye's view on what makes methods diffuse across various contexts. In WG4 both factors were taken into consideration. WG4 started by focusing on what factors in professional contexts of method users would play a role in method selection. To this end the work of Furniss (2008) was used. He had extensively studied contextual factors affecting usability evaluation methods in practice. For that he had taken three different perspectives. In one of WG4's workshops (step 1a), WG4 explored the applicability of those perspectives.

The contextual factors brought forward by the the perspectives were perceived as being relevant to WG4's aim, even though applying the perspectives as a whole proved to be problematic. A subsequent step in the process then was to create an overview of results from TwinTide's other three WG's and to try and combine them into an integrative model in a bottom-up manner. This led to an initial, tentative, static model describing how resource functions (WG1) and software quality attributes (WG3) related to each other and to methods, as well as to a dynamic model focusing on the situated process of method transfer (step 1b). This integrative model, however, did not prove to include the necessary types of contextual factors as they were identified in step 1a. A diffusion of Innovations (DoI) (Rogers, 1983) perspective was then introduced as a tentative integrative framework that was perceived as allowing integration of static, as well as dynamic process models, while at the same time it was thought to provide a view on the organizational context into which method transfer would be embedded. This became the framework approach adopted by WG4. A variant of the DoI method (Askarany, 2003) was used to consolidate the use of the framework for WG4's purposes, and to possibly suggest adaptations to it. Consolidation was done by conducting a survey questioning authors of reported real-life case studies collected for the TwinTide workshop at the CHI2013 conference (Law et al., 2013).

### 2.1.2 Multi-Criteria Decision Making Mechanisms

In selecting a method for a specific context many factors can potentially be involved. Numerous method attributes may have to be matched to numerous weighed contextual factors. Hence in the MoU studying Multi-Criteria Decision Making Mechanisms (MCDMs) had been included as one of WG4's tasks. Two expert members of TwinTide introduced the TwinTide consortium to MCDMs. Based on small-scale trials of MCDMs, using selected outcomes from the other WGs, TwinTide members were given the opportunity to get a feel for applying MCDMs; their applicability could be tried out empirically (see Figure 1, step 2). Finally, the use of MCDMs was positioned in the general framework for Method Diffusion across Contexts (step 3).

## 2.2 Process steps in detail

In the following sub-sections the process steps are described in more detail. The series of WG4 workshops are first presented, followed by Progress Step 2 (Fig. 1) MCDMs, and then Progress Steps 1a-1d (Fig.1), which lead to the description of the integrated framework in Section 3.

### 2.2.1 WG4 workshops

WG4 organized eight workshops, of which one was an open workshop organized in the context of the CHI conference (Law et al. 2013). In the following section the tasks for the workshops as well the main insights gained from them are summarized, with reference to appendices providing more insights into details.

**Table 1.** Overview of WG4 workshops.

| | Date | Place | Process steps |
|---|---|---|---|
| WS1 | November 2011 | Bertinoro, Italy | 0 |
| WS2 | March 2012 | London, United Kingdom | 0, 1a, 2 |
| WS3 | June 2012 | Skopje, FYR Macedonia | 1b, 2 |
| WS4 | September 2012 | Delft, The Netherlands | 1b |
| WS5 | October 2012 | Copenhagen, Denmark | 1c |
| WS6 | March 2013 | Coimbra, Portugal | 1c |
| WS7 | May 2013 | Paris, France | 1d |
| WS8 | June 2013 | Tallinn, Estonia | 1d |

### 2.2.2 Process step 0: Requirements for method selection tools

Work on exploring requirements and visions for method selection tools was done in workshops in Bertinoro (WS1) and London (WS2) (see appendix A: report of the Bertinoro workshop). In Bertinoro we brainstormed and discussed about who would be the users of a method selection tool and tried to imagine what they would be using the tool for. Identified potential users of such a tool included researchers, practitioners, educators, quality assurance managers, students, and project managers. The groups assumed that these people could use these tools for activities such as: learning about new methods, planning and monitoring projects, refreshing and updating knowledge about methods, documenting design and evaluation deliverables, and selecting methods in specific project contexts.

**Fig. 2.** Groups in the workshop in Bertinoro.

After this brainstorm, workshop participants were asked to come up with visions on what such a tool could be like. The aim of that part of the workshop was to make the experts think about the actual process of selecting a method for work to be done.

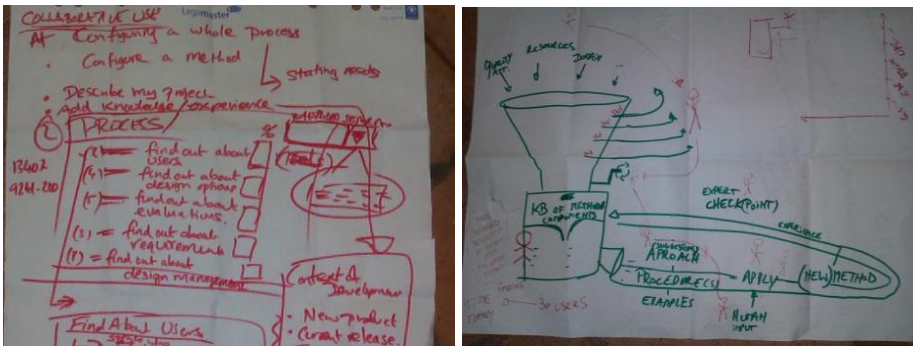



**Fig. 3.** Impressions of some visions on a method selection tool, resulting from the Bertinoro workshop.

In the London workshop (WS2) three existing online method selection tools were presented by people involved in the development of those tools[2]. At the time of the workshop, UCDToolbox was still under development. In the workshop participants were first given a summary of the method selection tool requirements as they were developed in the Bertinoro workshop. They were then invited to evaluate the existing tools with respect to these requirements. The aim was to determine to what extent these existing tools could probably serve as starting points for developing a tool that would facilitate method transfer, based on the insights gained in TwinTide thus far.

| | THE TRISTAN TOOL | USABILITY PLANNER | USEWELL |
|---|---|---|---|
| usable | ? | ? | ? |
| customization | ✓ | | |
| use cases/reports | ✓ | | ✓ |
| recommendations | | ✓ | |
| contributions | ✓ | | |
| collaboration | ✓.. | | |
| planning, learning | | | |
| configuration | | | |
| appropriateness | ✓ | | ✓ |
| DRE goals ID | | | |
| context ID | ✓ | ✓ | |

**Fig. 4.** Example of a worksheet evaluating three existing method selection tools with respect to WG4 requirements. 'The Tristan tool' refers to UCD Toolbox.

[2] Nigel Bevan presented www.usabilityplanner.org; David Geerts presented www.usewell.be; and Tristan Weevers presented www.ucdtoolbox.com

*Insights gained from taking the method selection tool perspective.*

The workshop helped to identify stakeholders other than actual method users, which could potentially benefit from or otherwise play a role in the application of method selection tools (e.g., planners and managers). In other words the workshop led to initial considerations about the social context that could play a role in adopting or not adopting specific methods. Whereas in discussions in the other WGs the focus had mainly been on matching method attributes to 'work to be done', the requirements that surfaced in Bertinoro and London implied that additional factors could play a role as well, such as awareness of the existence of methods, and knowledge about the specifics of applying a method (e.g., see appendix A: the activities for which the imagined tool would be used, such as 'learning about new methods' and 'updating or refreshing knowledge about specific methods').

The visions expressed in the Bertinoro workshop (WS1) depicted structures and flowcharts for method selection tools. The added perspective this brings is that transfer is not just about the question of whether there is a match between attributes of methods and attributes of an application context (static view), but that it also deals with questions such as 'what is the process of selecting a method like?' (dynamic view). Some of the visions expressed the view that a method selection process not necessarily is a straightforward top-down process of defining (abstract) criteria for candidate methods and then presenting a search result, but that it may also need to incorporate a more bottom-up approach. A bottom-up approach can be thought of as providing an overview of methods to narrow down, implicitly rather than explicitly using selection criteria. The London workshop (WS2) showed how the UCDToolbox seemed to fit this latter requirement best. However, it became clear that each of the tools would need substantial revisions to be able to provide the variety of different desired goals for which such a tool would be used, and that the number of parameters to be used for matching methods with contexts would need to be much larger both in terms of method parameters, as well as in the number of context parameters.

### 2.2.3 Process step 2: Multi-criteria decision-making mechanisms (MCDMs)

MCDMs are mechanisms that are specifically useful when having to make choices based on large numbers of parameters. WG4 worked on MCDMs in the workshops in London (WS2) where the MCDM mechanisms were introduced, in Skopje (WS3) where they were applied, and in Delft (WS4) where the results were presented.

In the London workshop one of the MCDM experts in TwinTide provided the workshop participants with an introduction to the topic. The history of such approaches was explained, a number of example mechanisms was briefly explained, as well as a classification of MCDMs. From the explanations it became clear that MCDM approaches would lend themselves both to top-down (objective/criteria-led) as well as to bottom-up (alternative-led) approaches.

In the Skopje workshop (WS3) the applicability of MCDM approaches to method transfer situations was explored. A group experiment was set up by another MCDM expert in TwinTide, with participants of the workshop (TwinTide members) providing the data and parameters to be entered in a MCDM procedure. For the experiment participants were provided with a transfer story (developed in WG3), a limited

number of method attribute types (based on WG1 and 2 work) (e.g., level of documentation of a method, development phase to which it applies, time needed), and six existing design & evaluation methods (e.g., heuristic evaluation, card sorting, eye tracking). In the following sections the concept of MCDM will be explained a more detailed, and the exercise that was done in Skopje is described in more detail including the results. This section is written by Paulo Melo and Bojan Sdrjevic

*Multi-criteria decision making concepts.*

Multi criteria decision-making (MCDM) is a useful instrument in handling structured decision-making problems. Structuring the decision-making process means to identify decision elements (DEs) and key players, and in turn decision-makers (DMs), then to define important decision lines. Fundamental to decision making is the understanding the problem in hand, following selected consistent and coherent methodology eligible for evaluating DEs, and preserving DMs to be responsible in validating objectives, criteria and alternatives (such as willingness, good attitude, best knowledge etc.), and finally, understanding that there will be consequences of decisions made. Decision making is a process that requires considering the spatial and temporal component of the problem, recognizing the process in the selected methodology of evaluating DEs and envisioning the necessity of feed-back loops.

Typical stages in the decision-making process are (not necessarily in that order and with usual overlaps and feedback): a) information collection and management; b) modelling and rational decision support; c) visualization and the human interface; d) group decision-making; e) knowledge capture and representation; and f) decision support system integration.

A few hints for implementation of a structured decision-making process are:

1. Carefully analyze the decision problem in hand and structure it.
2. Learn from the others.
3. Use existing software tools (develop additional or new only if necessary), and
4. Follow-up the consequences of the decision made, and be ready to repeat calculations.

The inherent characteristics of the group decision problem should determine the methodology to use and what outcome can be expected. For instance, in MCDM applications, the size of a group may become critical because most multi criteria methods rely on an assumption of homogeneity. Arguments can be made that in larger groups this assumption may be violated, and therefore a group clustering based on similarities in decision-making preferences, whenever the group is of an intermediate or large size, can be the best option. However, even in subgroups some of the problems related to a group may reappear, such as the tedious and time-consuming process of eliciting judgments, and additional sub-grouping could be necessary.

Some classes of commonly used methodologies and methods in the MCDM field include: 1) Utility based models, 2) Ideal point methods and 3) Robust techniques.

**Utility based models** are based on a (sometimes implicit) valuation of the utility of each alternative

- SAW (Simple Additive Weighting)
  Weighted sum of alternative evaluations
  Simple, but even so the results can be very close to more sophisticated methods.
- MAUT (Multi-attribute Utility Theory)
  An utility function (usually concave)is constructed and used to directly evaluate alternatives
- AHP
  Uses pair wise comparisons
  Assumes a linear additive utility function
- Outranking methods (e.g. PROMETHEE, ELECTRE)
  Based on partial orders (and implicit utility functions)
  Produce a (weak) ordering of alternatives

**Ideal point methods** are based on finding the best solution as function of a distance to the ideal (and sometime the anti-ideal) solution

- CP (Compromise Programming)
  These models are based on finding the best solution as function of a distance to the ideal (and sometime the anti-ideal) solution
- TOPSIS (Technique for Order Preference by Similarity to Ideal Solution)
  The most preferred alternative should not only have shortest distance from 'ideal' solution, but also longest distance from 'negative-ideal' solution.

**Robust techniques** assume that the knowledge required to build the decision making models may be incomplete or imprecise.

- Fuzzy Optimization, Grey Programming
  Assume that the problem is described using terms with some semantic ambiguity, and use fuzzy logic or grey logic to combine those terms
- Stochastic Programming (e.g. SMAA)
  Assume that the problem is defined as stochastic distributions rather than fixed numbers, and calculate expected outcomes

It should be noticed that not all of the previous models are commonly used in group decision making contexts (some require adaptations to be used). Although some versions to support this methods can be commercial, most are academic and readily available in the Internet.

*Applying MCDM to UX method selection*

To verify how MCDM could be applied to the choice of an UX method, we performed a practical experience to discover whether a group could cope with the informational requirements to apply MCDM on a method selection task (this work is described in detail in Melo and Jorge (2013), from which the following was copied with adaptations). This experiment was performed by a set of 20 UX experts, organized in four groups (O1 to O4). The exercise intended to provide an answer to which method would be "best suited" to a particular situation.

**A. The exercise context.**

In the actual exercise, we restricted our choice to six methods, chosen to provide a sufficiently diverse mix of characteristics:

- Heuristic Evaluation (Nielsen & Molich, 1990)
- Cognitive Walkthrough (Nielsen, 1994)
- UX curve (Kujala et al., 2011)
- Eye tracking (Yarbus, 1967)
- Card Sorting (Nielsen & Sano, 1995)
- Contextual Inquiry (Wixon et al, 1990)

And tried to select which one best suited this particular situation (i.e. story inspired by WG3 work):

- As a serious game designer I want to use collaboration techniques from e-learning applications to improve the effectiveness of my developed games.

The situation was deliberately stated in an open formulation, so that there could not be an "a priory" obvious solution to the group, and members could discuss among themselves what characteristics would be required in this kind of situation.

To assert fit between method and situation, the experiment was restricted to six method attributes/criteria (notice these criteria are shown as requirements for the method, but they could also be stated as resources available in the situation –these criteria were selected from a shortlist, provided by UX experts, of relevant method characteristics, but other criteria could be used instead; criteria were inspired by the work of WG1 and 2):

- Depth: Does the method go in-depth in its results (e.g., very detailed usability problems), or does it provide just some general hints or ideas?
- Documentation Level: Documentation of procedure, is it a method that is well documented and prescribes actions in much detail, like, for instance, GOMS (Card et al, 1986), or is it mainly a general approach (e.g., usability testing)?
- Structure: Level of structure/degree of formality. Some methods, e.g., GOMS are very structured and formal; Heuristic Evaluation is rather unstructured and informal.
- Time: How much time does it take to perform the method to its completion?
- Expertise: How much expertise is required to use the method? Some methods require a certain level of expertise in some domain (e.g., usability or application domain).
- Phase: When is the method best applied? Some methods are best suited for early design phases, other are for final products only, and other still for inspiration phases, etc.

As a previous step, and to familiarize themselves with the process, the groups were asked to create evaluation scales (in fact, define the evaluation domain) for each of the six characteristics. There was a need to use Likert-like items for some criteria,

since the groups could not agree on common descriptions to the method characteristics for the applicability of Likert scales to similar applications, see Norman (2010). Notice also that for most characteristics the groups could not agree on a "desirability" direction associated with the characteristic, which means that using it directly would be difficult.

### B. Collecting data.

The four groups evaluated both the methods and the situation according to the previous method attributes (in such a way that each method attribute would be evaluated by two different groups – this was made to both lessen the load on each individual group and also to mimic a decentralized application where not everyone knows all the methods or its characteristics. The raw data of choices made by the different groups was aggregated (e.g. when group elements defined several values for a particular criterion/method attribute, an average group value was used). Time values were converted to a common unit. This data was then organized in two "synthetic" groups, SG1 and SG2 each assembling some evaluations from three (of the four) original groups.

Converting the data to numerical inputs, by normalizing each value from each scale to a value in the 0-1 interval, allowed MCDM methods to be applied. Additional calculations, using Pearson correlations, has shown that the evaluations by the two groups on the different methods characteristics were usually similar, with 4 of the characteristics having correlation above 0.6. However, the two groups had a very distinct view on the situation P, which can be made visible by performing an additional Pearson correlation on the situation index column for SG1 and SG2. Doing so, we obtained $\rho$=-0.881, indicating that there is a strong negative correlation between the groups on the two situation indexes.

### C. Converting collected data to MCDM input

As a result of the previous findings, it was decided to handle each group separately, but to average the evaluations of the methods on the characteristics where those evaluations converged (in fact, all but Depth and Phase). This way, the shared knowledge was used to improve the evaluation of the methods on Documentation Level, Structure, Time and Expertise, but separate group evaluations were used for the Situation description and for the Depth and Phase characteristics.

Using the procedure described previously, the difference on positions stated by each group on the different criteria and on the different situation descriptions was computed, after converting those positions to a numerical value in the range [0,1]. With this calculation, a method that completely agrees with the situation on a particular criterion/method attribute would receive the value 0, whereas a method that is the opposite of the situation on that criterion would receive a value 1. Notice that the situation is seen as different among the two groups, and as such, even on "common" criteria, a method that is seen as "close" to the situation on one group may be seen as "away" from it by the other.

The evaluations diverged a little among the two groups. After applying the procedure described although the dispersion inside each evaluation (measured by the

standard deviation on the values by criteria/group) is not very high, it should be noticed that even groups that achieve similar average distance to the situation on a method attribute don't usually select the same value for the evaluations (that is, although the average distance achieved is similar, the distance differ on the individual methods).

The most extreme case for different evaluations is found on the "time" criteria, where while both groups give not completely different evaluations on methods, they differ in such a marked way on the situation description that for a group almost all methods are close to the situation in terms of time while for the other almost all methods are very different from its situation in terms of time.

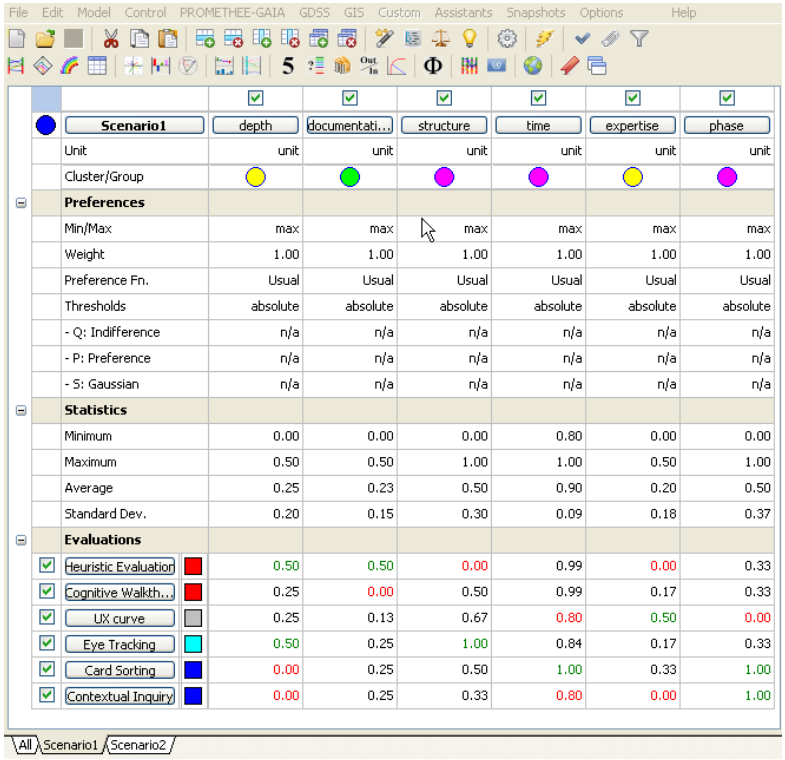
An almost as stark difference among groups is found on the "phase" criteria, where a group considers the method should be applied on a functioning product where the other considers it should be applied on a design concept (although in this case, the results of the differences are not as visible as on the "time" criterion since the short ranges considered limit the divergence).

Notice that the algorithmic approach chosen to handle the differences is usually considered inadequate if the groups are to be considered as handling the precise same problem. A more common way to handle the divergences would be to reconvene the decision makers to try to get them to achieve a common understanding of the situation (and of those criteria where marked differences were present). However, since this was impossible in the exercise conditions (and would also be very difficult to achieve in practical usage of this approach, where the "evaluators" of criteria and the "users" of the methods may never be in direct contact), the approach presented seems to satisfy the homogeneity of preferences requirement for MCDM procedure application.

### D. Applying the MCDM procedures

Once the data was computed according to the previous procedure, the actual MCDM process could be applied. For this exercise, two different MCDM methods were used, PROMETHEE (Brans et al, 1986) and SMAA-2 (Lahdelma et al, 2001). Those MCDM methods where chosen because of their fit to the data collected, being also simple enough to not create particular difficulty in the results interpretation, while powerful enough to support scenarios and output incomparability (PROMETHEE) and imprecise/stochastic input data (SMAA-2).

The values were then input into the Visual PROMETHEE software (Mareschal & de Smet, 2009), as shown in Figure 1. To support the different groups, the information was defined as two different scenarios.



| Scenario1 | depth | documentati... | structure | time | expertise | phase |
|---|---|---|---|---|---|---|
| Unit | unit | unit | unit | unit | unit | unit |
| Cluster/Group | | | | | | |
| **Preferences** | | | | | | |
| Min/Max | max | max | max | max | max | max |
| Weight | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| Preference Fn. | Usual | Usual | Usual | Usual | Usual | Usual |
| Thresholds | absolute | absolute | absolute | absolute | absolute | absolute |
| - Q: Indifference | n/a | n/a | n/a | n/a | n/a | n/a |
| - P: Preference | n/a | n/a | n/a | n/a | n/a | n/a |
| - S: Gaussian | n/a | n/a | n/a | n/a | n/a | n/a |
| **Statistics** | | | | | | |
| Minimum | 0.00 | 0.00 | 0.00 | 0.80 | 0.00 | 0.00 |
| Maximum | 0.50 | 0.50 | 1.00 | 1.00 | 0.50 | 1.00 |
| Average | 0.25 | 0.23 | 0.50 | 0.90 | 0.20 | 0.50 |
| Standard Dev. | 0.20 | 0.15 | 0.30 | 0.09 | 0.18 | 0.37 |
| **Evaluations** | | | | | | |
| Heuristic Evaluation | 0.50 | 0.50 | 0.00 | 0.99 | 0.00 | 0.33 |
| Cognitive Walkth... | 0.25 | 0.00 | 0.50 | 0.99 | 0.17 | 0.33 |
| UX curve | 0.25 | 0.13 | 0.67 | 0.80 | 0.50 | 0.00 |
| Eye Tracking | 0.50 | 0.25 | 1.00 | 0.84 | 0.17 | 0.33 |
| Card Sorting | 0.00 | 0.25 | 0.50 | 1.00 | 0.33 | 1.00 |
| Contextual Inquiry | 0.00 | 0.25 | 0.33 | 0.80 | 0.00 | 1.00 |

**Fig. 5.** Data used for PROMETHEE

With this data, PROMETHEE was used to assess the understanding from each group regarding which methods could be better applied to the situation. Assuming all method attributes are considered as equally important (equal weights on the method characteristics in Figure 5) PROMETHEE would give us the result present on Figure 6 using the PROMETHEE Diamond, see Mareschal & de Smet (2009), presenting simultaneously two partial rankings for the methods (the higher the methods are in the ranking the best the fit between method and situation). In this picture is easy to understand that two methods are considered to be a better fit (Card Sorting and Eye Tracking) while the rest are seen collectively as somewhat similar among themselves and an overall worse fit for the situation (with Contextual Inquiry as the worse fit).Although it could be claimed that given the results shown Card Sorting should be seen as a better fit than Eye Tracking, PROMETHEE by itself cannot give us an absolute ranking between the two, but can say that, giving equal importance to all method attributes, those two are certainly a better fit than the rest.

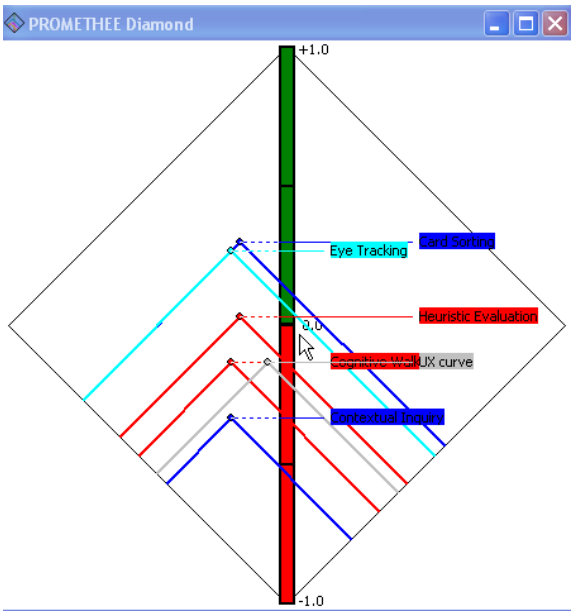



**Fig. 6.** PROMETHEE diamond (SG1)

It should be noticed that the rankings presented are the combination of fit among different attributes, and there is some compensation among the different evaluations to reach the final rankings.

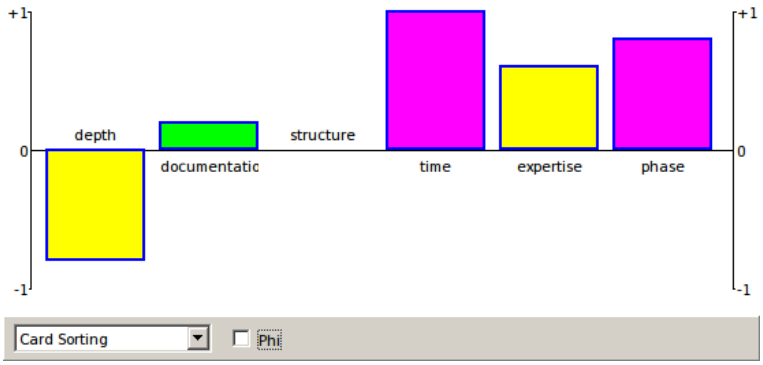



**Fig. 7.** Fit by method characteristics for card sorting (SG1)

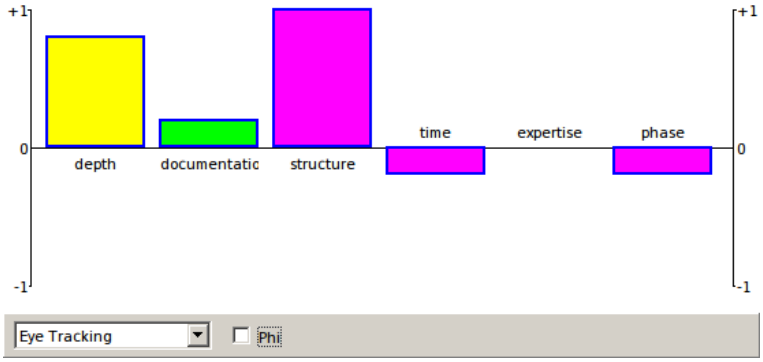



**Fig. 8.** Fit by method characteristic for eye tracking (SG1)

As seen in Figure 7, Card Sorting is shown as having better fit on time, expertise and phase and worse fit on depth (with positive fit on documentation), whereas Eye Tracking (Figure 8) has an very good fit on structure and a little smaller fit on depth, and somewhat negative fits on time and phase. Overall those fits add up (using equal weights) to results that are broadly similar. Some other methods (like Heuristic Evaluation) can have worse fit on more method attributes and smaller fit with the remaining, which explains their relative ranking.

Since the overall ranking is dependent on all the method attributes, a few good results may overcome extremely bad results on a particular method attribute. PROMETHEE supports the concept of veto to override this effect, but the experiment didn't use this possibility. Therefore, the results shown may overstate the fit of methods which would be disqualified by such veto.

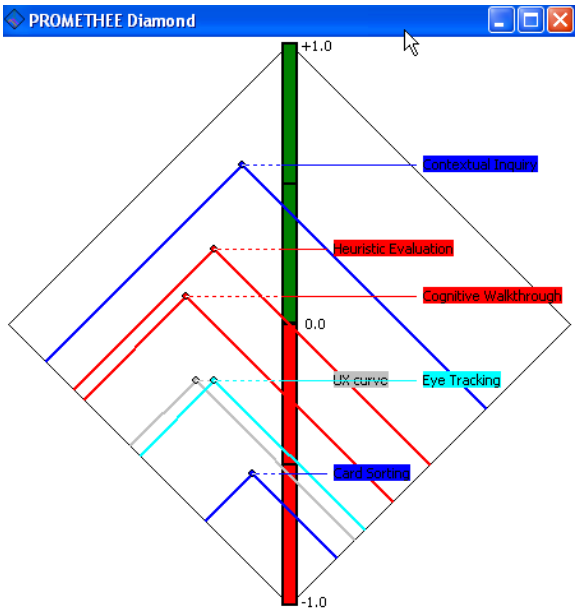



**Fig. 9.** PROMETHEE diamond (SG2)

However, applying the same procedure as before to the data from SG2 will give us a very different ranking of methods (as seen in Figure 9). In fact even if we just use the PROMETHEE II ranking, it is easy to show that the two groups have completely different opinions regarding the method which is best suited for the situation (see Figure 10).

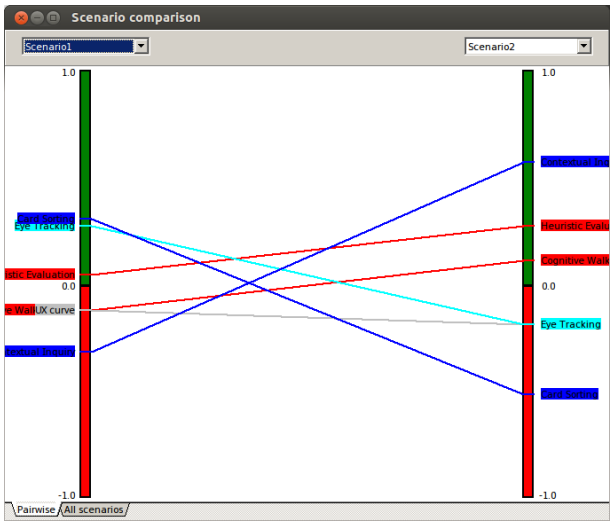



**Fig. 10.** Group rankings comparison

Notice however that these results mimic strongly the original negative correlation found in the situation description among the two groups. Since the situation perceived by those groups is mostly symmetrical, it isn't very surprising that a method that ranks best for a situation is ranked last in the other.

The results also depend on the relative importance given to method characteristics, but agreeing on importance to give each criteria can be hard. Some MCDM procedures can be used to provide a statistical distribution of results without fixing "a-priori" its importance (the MCDM procedures which do so are often called preference disaggregation approaches, see Siskos et al (2005)). This exercise was performed using the SMAA-2, the stochastic multi-criteria acceptability analysis procedure (Lahdelma et al., 2001). In this, each alternative is assigned a probability of belonging to a rank (from the 1, the best, to 6, the last one) in terms of weight combinations. Using the previous data, in almost all of the possible weight combinations "Card Sorting" would be selected as the best fit, and in over 95% "Contextual Inquiry" is the worst (although Eye Tracking is not always seen as the second best option). This shows a very strong agreement with the previous results, achieved using a different MCDA procedure and a different approach.

Notice however that these results were arrived at using the average group evaluation for a few characteristics (achieved by averaging the evaluations of the pair of synthetic groups). To try to check whether such averaging distorted the results the experiment were remade in SMAA-2 using as parameters not exact values but stochastic values with normal distributions with the same average and standard deviation that were achieved by averaging the groups. As one can see (Figure 11) Contextual Inquiry and Card Sorting are still overwhelmingly considered the worse and best fit, but there is now about 15% of evaluations where they don't have those ranks. But while some results don't agree so impressively with the previous ones, other are even more akin to the ones obtained with PROMETHEE, namely the classification of Eye Tracking. This reinforces the overall feeling of robustness for the results found.

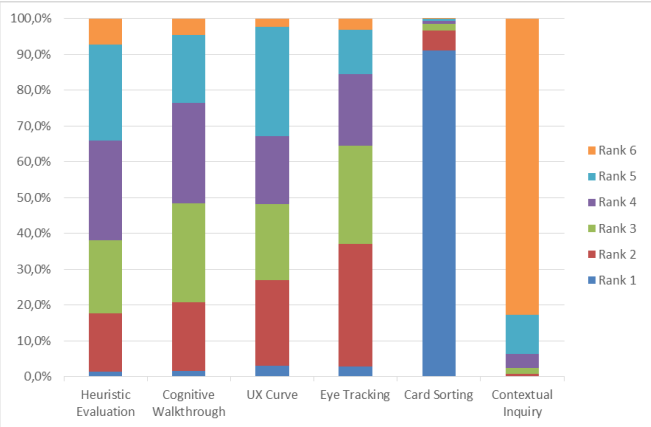



**Fig. 11.** SMAA rank acceptability indexes – using stochastic indexes information (SG1)

*Some insights gained from exploring the M-C mechanisms.*

From the exercises it was clear that MCDMs could be fruitfully used for identifying a best fit of methods for a situation. A precondition for a good fit, however, would be that the data provided about method attributes (types of variables, and how methods score on them), as well as about contextual attributes (how does the context score on the variables) is reliable. Even though most participants in the workshop can be considered experts on design and evaluation methods, it proved to be very difficult to reach consensus at various points. There was disparity between the various participant groups on how to understand the various method attributes in terms of how (using which variables) they can be measured. Expertise with respect to the qualities of the various methods seemed to be reasonably consistent, but evaluation of the story in terms of variables proved to be highly unreliable. The latter is partly understandable, because the story didn't provide much context. However, consequences for a method selection tool would be that either the people that fill the database inform themselves about the characteristics of a huge variety of possible contexts, or the tool would require users to first fill in a large amount of (more or less abstracted) data about the context for which they would like to use a method. Filling such data proved to be a tedious job.

### 2.2.4 Process step 1a: Exploring contextual factors in professional practices

In his doctoral thesis *Beyond Problem Identification: Valuing methods in a 'system of usability practice'* Furniss (2008) reports about his study on factors that influence the choice of usability evaluation methods in the professional practice of website design. Furniss stresses 'the importance of contextual factors and the need for system descriptions' because 'Usability Evaluation Method (UEM) adoption and adaptation cannot be fully understood devoid of context'. In analyzing the contextual factors that may be of influence to the choice of methods he explores three different perspectives: PMO, Distributed Cognition and Resilience Engineering. As usability evaluation is one of the approaches relevant to TwinTide, it was assumed that contextual factors that are important for applying Usability Evaluation Methods may also be relevant to other Design and Evaluation Methods considered in TwinTide. The perspectives were explained to the participants in the London workshop (WS2) as follows:

- **PMO perspective.** PMO stands for Project planning, Method adoption and adaptation, and Output of method. In this perspective a project is metaphorically seen as a flow from P to M to O through a landscape of contextual factors. Contextual factors are then categorized into: 1) technical factors (which include the Human Factors (HF)/usability issue at hand), 2) social factors (which include personal preferences and relationships), 3) structural factors (which include the stage and organisation of the project), 4) communication factors (which include informal and formal style of reporting), and 5) resources (which include times, budgets and capabilities).

- **Distributed Cognition perspective.** The Distributed Cognition perspective uses an information processing metaphor vocabulary to describe systems and considers

the wider factors that functionally affect the information flow, e.g. how social structures, use of tools and artefacts, procedures and changes over time influence UEM practice. This perspective helps in considering how methods are affected by a number of factors in the system into which it is embedded. Identified factors include: 1) social aspects (including power relationships), 2) information flow (including the timing and type of reporting mechanisms), 3) artefacts (including capability and availability of tools), 4) evolutionary aspects (including the changes of methods and tools over time), and 5) physical aspects (including the closeness of the interaction between the HF/usability practitioner and the client group).

- **Resilience Engineering perspective.** The Resilience Engineering (RE) perspective highlights factors influencing the performance of HF/usability practice in a functional manner. It emphasizes the variances that happen in normal HF/usability work and how practitioners use their expertise to adapt their methods and actions to fit the context. From this perspective we can see how factors categorized in seven RE themes may influence method use: 1) goal conflicts between practitioner and client: efficiency-thoroughness trade-offs in terms of the depth of work and research performed, 2) values: balancing different goals to make useful and pragmatic contributions to their clients, 3) adaptation of methods in terms of accounting for the normal variances on project work, 4) reflecting in practice to respond to variances, 5) expertise in terms of recognizing patterns and knowing what adaptations are suitable, 6) sharp-end/blunt-end distinction to highlight the influence of the client context, practitioner context and academic context; there are levels of contextual influences ranging from very close-by (e.g. local workplace factors) to more distant and abstract factors (such as moral and social), and 7) tight and loose coupling relating to the level of rigidity in adhering to deadlines and budgets and the adaptations that are made to method prescriptions.

The aim of this part of the process was to explore the relevance of the contextual factors brought about by the three perspectives, and to find out to what extent there would be clear relations between types of contextual factors (as identified in the three perspectives) and categories of method attributes (i.e., resource functions). If more or less stable relationships between these two could be identified, this would help in selecting methods. For example, the category of axiological resources functions is about the values underlying a specific method. Suppose one searches for a method to be used in a work context where low costs are valued over high quality, one could then pre-select candidate methods based on axiological resource functions relating to cost and quality and one wouldn't have to dive into the details of many other resource functions relevant to the method.

In the workshop the group of participants were split up in three groups, each group dealing with one of the three perspectives. Participants were given a list of method attributes categorized according to the resource functions identified by WG1 and the categories identified by WG 2 and 3. Participants were asked to stick Post-It notes, with one method attribute each, on flip charts showing contextual factors related to the perspective they were dealing with (i.e., PMO, Distributed Cognition or

Resilience Engineering). See appendix B for the slides introducing the exercise for this workshop.


**Fig. 12.** Example of categorization of method attributes/ resource functions according factors in the Distributed Cognition perspective. The pink one at the top right are method attributes that did not fit any of the categories.

Although in the end most of the method attributes/resource functions were connected to factors relating to the perspectives, the participants felt that they could not identify stable relations between the two which would hold across contexts. Such relations would be too situated. However, it was acknowledged that in itself the factors as indicated by the perspectives would be relevant in practice. It was concluded that none of the perspectives provided a framework that would be comprehensive enough to deal with method attributes in relation to all possible types of factors of a work context. A different perspective would be needed for that, but factors similar to those identified by the perspectives would most likely have to be part of that framework.

### 2.2.5 Process step 1b: Bottom-up modeling of input from WGs 1, 2 and 3

After the work on the MCDM experiment in Skopje (WS3) a diagram was presented in a first attempt to devise a framework integrating results from the other WG's (see Figure 1). The diagram consisted of a static model showing relations between various elements (see Figure 13).

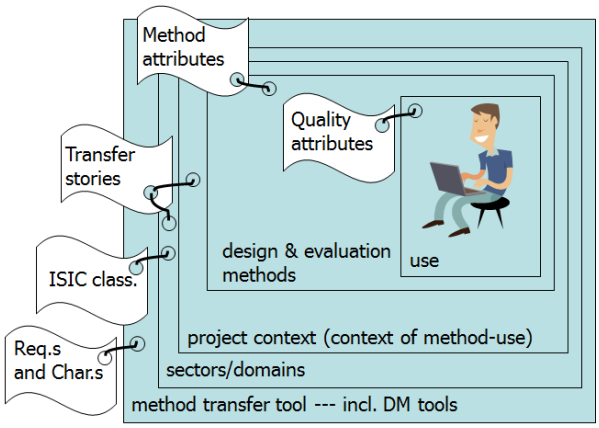



**Fig. 13.** First attempt to combine insights from the various workgroups into a static model.

Again the need for a complementary process view on transfer was expressed, and it was decided to address this in the subsequent workshop in Delft (WS4). Moreover, it was decided that until now WG4 had drawn upon their own expertise and imagination for coming up with the framework. Therefore, in Skopje (WS3) the decision was taken to also collect real-life cases of situations in which new methods were introduced or of method transfer. It was decided to do this by organizing an open workshop at the CHI conference in April/May 2013 in Paris, France (Law et al, 2013).

In the Delft workshop (WS4), WG4 started considering how a static model and a dynamic model could be integrated. Figure 14 shows an example of one of the attempts.

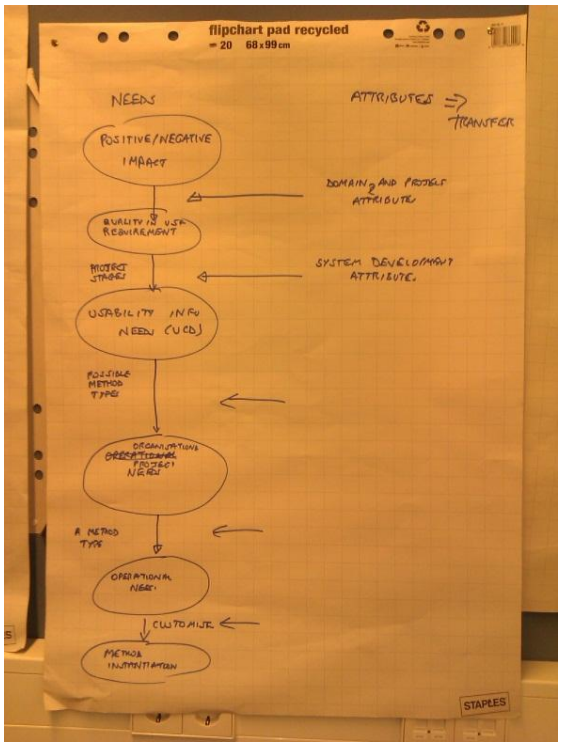



**Fig. 14.** One of the initial attempts in the Delft workshop (WS4) towards a combined static and dynamic model.

### 2.2.6 Step 1c: Introducing the Diffusion of Innovation (DoI) perspective

The next step in WG4's process was to try and converge towards one framework. In the Copenhagen workshop (WS5) participants were presented with Rogers' (1983) Diffusion of Innovations framework for the first time (see next section) and were asked to explore to what extent its concepts made sense to the participants. In the workshop in Coimbra (WS6) this was done in a more systematic manner. Also, the process model from WG2, and the quality attributes from WG3 were included in that exercise to see to what extent these could be part of WG4's framework. See appendix C for the materials that were handed out. The main task for the participants was to find a way to combine these elements into a single framework. Obviously, this didn't lead to a combined framework during the workshop, but it provided some interesting insights. WG2's process model was the main focus of the discussion. Comments on that model included:

- It may be too linear
- If there is a method champion who actively pushes a method, than there is no such systematic selection process
- The model is not a real-life model, as real-life is much more chaotic
- In many cases there is not such strict procedure

- Sometimes methods are just chosen because they appeal to the client, how does that get a place in it?
- This is a useful selection model, only the implementation part should be extended.

For WG4 this meant that although this process model may fit certain situations in real-life, in a general framework such as intended for WG4, the process would have to be less prescriptive and more open.

In the next section a first attempt is presented to translate the Diffusion of Innovation approach to method selection and transfer. This was presented during the workshop (WS7) in Paris (Law et al. 2013). This section comes largely from the paper written for that workshop (Vermeeren & Cockton, 2013)

*Transfer of HCI methods seen as Diffusion of Innovation.*

Rogers (1983) defines *an innovation* as "an idea, practice, or object that is perceived as new by an individual or other unit of adoption." *Diffusion* is defined "as the process by which (1) an innovation (2) is communicated through certain channels (3) over time (4) among the members of a social system." Adopting the Diffusion of Innovation perspectives for TwinTide WG4s aims implies treating HCI methods or approaches that are applied in a new context as innovations.

Applying the diffusion of innovations framework means that the new context is seen from the perspective of a *social system*, and that next to social context factors, *communication* is seen as playing an important role in the adoption and implementation of new methods. Furthermore, the diffusion of innovations approach implies a *process view* of adoption, adaptation and implementation, rather than a static view on matching characteristics of a context to attributes of a method.

HCI method selection tools (e.g., UsabilityPlanner.org, UCDtoolbox.com, AllaboutUX.org) provide their users with assistance in finding appropriate methods or approaches for specific contexts. They do this largely based on matching method attributes to (presumed) attributes of the target context in which they will be used. Such tools may also provide advice on how to adapt methods to the situations. This was also the original approach taken in the EU Cost Action IC0904 TwinTide project [http://TwinTide.org], as well as (implicitly) in the EU COST action 294 MAUSE [http://cost294.org]. In the TwinTide project there has been a shift towards a focus on the *process of transfer*. The framework we propose here attempts to connect the two approaches, via *social system* and *communication* perspectives.

On TwinTide's WG1 resource functions were highlighted that play a role in determining if an approach can be used in a specific context and what needs to be modified or added to how an approach has been implemented in a preceding context. Part of the process of implementing an approach in a new situation is trying to match the various resources as objectively as possible, in order to find an appropriate fit for the work to be done. Insights from diffusion of innovation research add a further perspective. This perspective makes clear that even if there seems to be a perfect fit, there are other factors that play a role in deciding on an approach or on how to implement an approach. This perspective relates to the social context in which practitioner work. Below we will discuss three groups of findings from diffusion of

research that seem relevant to our case: *adoption-relevant attributes of innovations*, *change agent success factors*, and the *innovation decision process*.

### A. Adoption-relevant attributes of innovations.

Rogers (1983) mentions five main attributes of innovations that play a role in whether an innovation will be adopted in a social context or not. In our cases, when practitioners start working in a new (social) context, wanting to apply an approach they are familiar with, a similar situation may occur: not only should the practitioners themselves find a match between the approach and the work to be done, they will also be confronted with some social context in which they work. To this social context, the new approach may be an innovation, and usually they may have to modify an approach to increase the chance of a successful implementation. According to Rogers (1983) the following five attributes of an innovation (here: approach) as perceived by the members of a social system may play a role in the adoption process: 1) *relative advantage*, 2) *compatibility*, 3) *complexity*, 4) *trialability* and 5) *observability*.

*Relative advantage* is the degree to which an innovation is perceived as being better than the idea it supersedes (Rogers, 1983). When practitioners introduce a new approach in their social environment, it has a better chance of being adopted if the people in that social environment *perceive* the approach as having a relative advantage. Note that the word *perceive* is as crucial as *relative advantage* here (as it is with the following four attributes). If the practitioner's environment doesn't see the relative advantage there is a higher chance that they will resist the change in their usual way of doing things. *Compatibility* is the degree to which an innovation is perceived as consistent with the existing values, past experiences, and needs of potential adopters. For HCI approaches as innovations, this is largely related to the match of resources discussed in part 1, however, this attribute emphasizes that what matters, is how social contexts shape perceptions of matches. *Complexity* is the degree to which an innovation is perceived as relatively difficult to understand and to use. If the social context thinks a new approach is difficult to use, or if they don't understand it, this lessens the chance of them agreeing about using it. *Trialibility* is the degree to which an innovation may be experimented with on a limited basis. So if an approach can be tried out on a limited scale without too many risks, this helps in introducing it. *Observability* is the degree to which the results of an innovation are visible to others. The results of some ideas are easily observed and communicated to others, whereas some innovations are difficult to describe to others. The same is true for new approaches. If after using a new approach it is difficult to observe or describe whether there is any difference in results or not, this lessens the chance of an approach being adopted.

### B. Change agent success factors.

Much of the diffusion of innovations research is about the role of change agents. About change agents Rogers states: "A change agent is an individual who influences clients' innovation decisions in a direction deemed desirable by a change agency. In most cases a change agent seeks to secure the adoption of new ideas, but he or she may also attempt to slow the diffusion process and prevent the adoption of certain

innovations." Rogers concludes that "a change agent's relative success in securing adoption of innovations is positively related to 8 factors: (1) the extent of change agent effort in contacting clients, (2) a client-orientation, rather than a change agency-orientation, (3) the degree to which the diffusion program is compatible with clients' needs, (4) the change agent's empathy with clients, (5) his or her homophily with clients (homophily is compatibility as the degree to which pairs of individuals who interact are similar in certain attributes, such as beliefs, education, social status, and the like (Rogers, 1983)), (6) credibility in the clients' eyes, (7) the extent to which he or she works through opinion leaders, and (8) increasing clients' ability to evaluate innovations." Considering the situation of a practitioner entering a new context some of these may be considered relevant as well. Effort in contacting the client (1) doesn't seem to be relevant here, as we assume that the practitioner is in the same team. This would also mean that client-orientation and change agent's orientation (2) will generally be the same. Furthermore, the situations we consider do not deal with diffusion programs (3) deliberately aimed at spreading certain practices just for the sake of spreading them. Increasing a client's ability to evaluate innovations (8) comes down to change agents seeking to raise the clients' technical competence and ability to evaluate potential innovations themselves. This is a long-range endeavour, which is also not relevant to the cases we consider here. What remains are empathy (4), homophily (5), credibility (6) and opinion leaders (7). For HCI practitioners wanting to introduce new approaches into a new context, this means that this will be more easy if the practitioner *shows empathy with other team members, is more homophilous with them*, if other team members *see the practitioner as credible*, and if the change agent can refer to *other teams or people that use the practitioner's approach* and who are seen by the team as *opinion leaders*.

### C. The innovation decision process.

Rogers (1983) defines the innovation-decision process as "the process through which an individual (or other decision-making unit) passes from first knowledge of an innovation, to forming an attitude toward the innovation, to a decision to adopt or reject, to implementation of the new idea, and to confirmation of this decision." Rogers originally distinguished five stages in the innovation decision process (Rogers, 1983). Although he consistently talks about an individual *or other decision making unit*, these stages seem to relate to individuals making choices mostly. These stages are 1) the (awareness-) *knowledge* stage when the individual (or other decision making unit) is exposed to the innovation's existence and gains some understanding of how it functions, 2) the *persuasion* stage in which one may become interested in the innovation and starts forming a favorable or unfavorable attitude towards it, 3) the *decision* stage when activities are undertaken that lead to adopting or rejecting the innovation, 4) the *implementation* stage in which an innovation is put into use, and 5) the *confirmation* stage when an individual (or other decision making unit) seeks reinforcement of an innovation-decision already made, but he or she may reverse this previous decision if exposed to conflicting messages about the innovation. Evidence for a very clear distinction between implementation and confirmation stage is weak according to Rogers (1983). Rogers also discusses the innovation process in

organizations and in that he distinguishes the following stages: 1) *agenda setting stage* in which an organization becomes aware of a problem in the organization that needs to be solved or is confronted with an innovation that uncovers a thus far unknown need, 2) *matching stage*, in which an organization is trying to figure out whether it seems worthwhile to adopt the innovation or not and tries to imagine the consequences of the innovation when implemented in the organization, 3) *redefining/ restructuring stage*, in which a solution is sought for an imperfect match between innovation and organization, either by re-inventing the innovation or by restructuring the organization, and 4) the *routinizing stage* in which the innovation becomes part of the daily life.

Transfer is thus prepared for at the agenda setting stage and then achieved via the others. If we translate this to the case of the practitioner wanting to introduce an approach to and in a new context we could summarize the process as follows:

The practitioner in the new context makes the others aware of a candidate approach or of an organizational need (knowledge stage), and makes the organization aware that a certain approach could fit an organizational need (agenda setting). To be able to apply the new approach, the practitioner needs to persuade those most directly involved in applying the approach or at least get them interested to cooperate (persuasion). For the practitioner's work to be done, he or she would need to evaluate the match of approach resources to the new situation (decision stage) and for the organization it would mean matching how it would fit the organization: what is the effect on the organization, how does it benefit the organization (matching stage and decision stage). Once the decision is taken to start using the approach the implementation phase starts, involving actually redefining the method by selecting appropriate resources and at the same time restructuring the organization (redefining/restructuring stage). Once taken into use the routinizing stage and/or confirmation stage can start.

*Towards a framework for Diffusion of Methods*

The items discussed so far are summarized below and they are hypothetically related to the situation of method selection or adoption of methods. In this way WG4 explored to what extent the framework would have potential as the general framework WG4 was looking for.

### A. Resource functions and innovation process stages.

In case of rejection of a (proposed) method, or of having to adapt it, this may occur at different stages of individual (I) or organizational (O) innovation decision processes:

- Knowledge/awareness (I1 - i.e., people in the new context not being aware of the method or not knowing what it can do);
- Persuasion/interest (I2 - i.e., difficult to get people in the new context interested or to make them form a favourable opinion about the method);
- Agenda setting (O1 - i.e., difficult to convince people that using the method leads to fulfilling organizational needs).

In the above stages an initial match of resources is usually being made for axiological or scoping resources. Possible reasons for not adopting an approach at these stages are:

- Scoping: the method does not fit the purpose of the work or the usage/process context well enough;
- Axiological: the method takes a different perspective on the work than is desired in the new context (e.g., with respect to what is valued), or there are ethical problems in using the method.

Reasons for not using an approach or for having to adapt it can also be found in the following stages, in which actual decisions are being made and implemented:

- Decision/matching (I3/O2 - i.e., difficulties in the actual process of taking the decision on whether to start using a method or not; evaluating its pros and cons; thinking through the consequences of implementation);
- Implementation/restructuring/redefining (I4/O3 - i.e., identified mismatches between resources and work context that lead to adaptations or modifications of the method's resources; or to changes in the organization to make it work);
- Confirmation/routinizing (I5/O4 - i.e., problems in sustaining a method's use).

In the above stages, considerations concerning the following resource functions play a major role. This is especially so in the decision and matching stage. However, in the later stage they continue to play a role:

- Harvesting: the instrumentation or type of data that the method works with, does not provide the kind of data or insights that the new context (wants to) work with;
- Directive: there is something about the procedures in using the method that does not fit the new context, or the procedures are perceived as being too complex or as having a poor cost-benefit ratio;
- Expressive: the kind of output the method gives or the way important elements are expressed with the method does not match the expectations and/or standards for communication in the new context.

**B. Attributes of innovations.**

Rejecting a (proposed) method or having to adapt it often relates to attributes that are typical for innovations in general:

- Relative advantage: not enough relative advantage, or relative advantage not being clear enough;
- Compatibility: perceived problems of applying the 'old' method in the new context (i.e., team perceives a mismatch between resources and work to be done);
- Complexity: method perceived as being too complex to use, or too difficult to learn;
- Trialibility: method cannot be tried before deciding to use it;
- Observability: merits of the method are difficult to observe by people not directly involved in using it.

**C. Personal (Change Agent) Factors.**

Sometimes application of a specific method also largely depends on personal relationships. The following change agent factors can obstruct success for someone wanting to introduce a change (e.g., a new approach):

- Empathy: not enough empathy between the practitioner and the new team.
- Homophily: difficulties in identifying with and associating with the people involved in using the new method in the new context, making them feel they are on different wave lengths.
- Credibility: the other people in the new team just didn't believe enough of the presented benefits of using the method.
- Opinion leaders: there was a lack of opinion leaders (in the eyes of the other people involved) who are also in favour of using this method.

### 2.2.7 Step 1d: Consolidating the DoI framework based on real-life cases

An open workshop was organized at the CHI 2013 conference (WS7; Law et al, 2013) with the aims to "bring together HCI professionals who have method transfer experience and knowledge to share, analyze and synthesize insights so gained." (Law et al, 2013) The goal was "to collect well-structured case studies of professional HCI practices for constructing applied knowledge for adapting and combining resources of sets of methods to deal with contextual constraints". (Law et al., 2013). The workshop was organized by three of TwinTide's WG coordinators, the Chair and another TwinTide member.

Eighteen contributions were accepted in the workshop. They are categorized into three main groups (Law et al. 2013):

1. Case studies on work-oriented applications

   - Wardlaw, Cox and Haklay on health care systems
   - Gasik and Lamas on meeting room booking services
   - Ikonen and colleagues on wellbeing and recovery management
   - Hvannberg on accessibility and crisis management systems
   - Sikorski on the customer relationship management system for a call centre
   - De Guzman on employee profile management

2. Case studies on leisure-oriented applications

   - Sutcliffe and Hart on art galleries
   - Derboven on multi-touch interaction table
   - Lárusdóttir on multiplayer online games
   - Arhippaninen et al on 3D virtual music
   - Johnson on virtual hotel games
   - Väätäjä on city life exploration

3. Methodological /theoretical frameworks

   - Rantavuo and Roto on applying heuristic evaluation to study user experience

- Vermeeren and Cockton on analysing key concepts of the diffusion of innovation framework
- Cockton on the diffusion of the novel method worth maps
- Bevan on usability maturity assessment and process improvement
- Jokela on the practical value of interviews
- Springett and Law on the possible integration of appraisal theory and action cycle

In the Tallinn workshop (WS8) a slightly adapted DoI framework largely based on the work of Askarany (2003) was presented. Askarany's framework was developed based on studies about diffusion of advanced (cost and accounting) techniques in organizations. This made WG4 consider it to be a useful alternative to Rogers'(1983) DOI framework. For WG4 the main difference with Rogers' framework is the addition to the framework of 'Attributes of Adopters' as factor influencing adoption (see Figure 15). Another addition, less relevant to TwinTide is the addition of the element 'Generation of Innovations'.

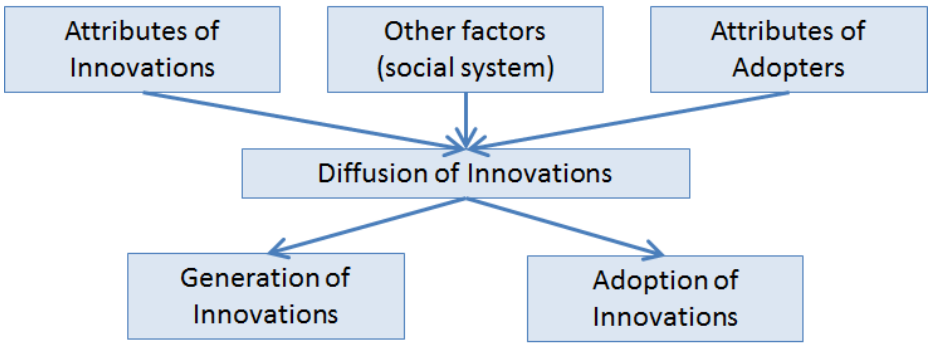



**Fig. 15.** Askarany's Diffusion of Innovation model (adapted from [1]).

In the Tallinn workshop (WS8) a tailored version of this framework was used as a pilot study in a small group exercise on some of the cases described in the CHI workshop papers. Participants were asked to read through one or two of the given papers and search for phrases in the case descriptions that seemed to refer to elements of the presented framework. Table 2 shows the framework elements as presented to the participants in the Tallinn workshop (WS8).

**Table 2.** Diffusion of Innovation framework elements as they were used in the Tallinn workshop (WS8) (adapted from Askarany, 2003)

| | |
|---|---|
| **Stages of process**<br>• Agenda setting:<br>– Awareness (of need or of approach)<br>– Attitude forming/getting interested<br>• Considering match between resources and context<br>– Decision making<br>• Implementation<br>– Redefine approach/method<br>– Restructure organizational context<br><br>**Attributes of adopters**<br><br>• Organizational structure, - culture and strategy<br>– Size of organizations,<br>– Aggressiveness and innovativeness of managers,<br>– Level of information adopters have about existence and characteristics of an innovation,<br>– Resistance to change,<br>– Technical skills of users of an innovation,<br>– Competition (wanting to gain an advantage over competition),<br>– Availability of resources in organization<br><br>**Attributes of social system:**<br><br>• Social system: Social structure, norms, opinion leaders and change agents in a social system<br>• Communication channels through which new ideas are communicated | **Attributes of innovation:**<br>• Relative advantage:<br>– Social prestige, savings in time and effort, low initial cost, degree of economic profitability, increase in comfort/decrease in discomfort, immediacy of output (some may not be apparent prior to implementation), flexibility or capability of modification of an innovation<br>• Compatibility:<br>– Degree of consistency of innovation with needs, expected values and norms of potential adopters and their social systems<br>• Complexity:<br>– Degree to which an innovation is difficult to understand and use<br>• Trialability:<br>– Degree to which an innovation can be tried on a limited basis before full implementation<br>• Observability:<br>– Degree to which results of an innovation can be observed by potential adopters |

The exercise done in Tallinn (WS8), together with comments from some of the authors of described papers who were present at the workshop gave confidence that the framework could well be used to provide a systematic overview of the factors influencing method selection. However, it also became clear that the written case descriptions only, would provide too little information. Hence it was decided to ask the authors of the papers for further information about their cases by means of a survey, structured around the framework elements. The survey was setup in collaboration by two WG-leaders and two other TwinTide members.

*Survey for consolidating the framework.*

Tables 3, 4 and 5 present the survey questions (the full text and templates can be found in Appendix D). The survey was sent to the authors as a Ms-Word template, on July 22 2013, and authors were asked to respond by August 22 ultimately.

**Table 3.** Part 1 of the survey on DoI factors.

**What do we ask you to do**

We ask you to go back to what you remember about the described case as it happened in reality.

We ask you to indicate for each of the items in the form below, if you think the described item played a role:

a) in deciding to use the method/approach described in your paper, and/or

b) in deciding to NOT use a possible alternative for the method/approach described in your paper.

If there is any text (sentence or phrase) in your workshop paper that supports or clarifies your answer, then please copy-paste that in the form below. If there is no such text in your paper, but you remember that something relevant happened, then please write a short phrase about that (you can keep it very brief). In case any of the questions may not be applicable to your case, please indicate so in the relevant field.

**1. Adopting a new method/approach**

What can you say about introducing the method(s) described in your paper?

*Awareness: How did awareness of existence of the method/approach affect considering it in your team or organization? E.g. was it neglected earlier because of the lack of awareness?*

*Attitude forming: How did attitudes in (or outside) your team or organization affect considering the method/approach for use in your organization? E.g. was it neglected earlier because of the attitude of some people?*

*Other: Were there other aspects that influenced the initial phase of taking the method/approach into consideration in your team or organization?*

**Decision making**

What factors in relation to this decision making process do you remember were crucial for adopting your described method?

*Method characteristics matching the work: What aspects of the method were the key influencers in taking the decision of using (or not using) it?*

**Implementation**

What adaptations to methods or to work contexts do you remember played an important role in the successful use of the method/approach?

*Redefinition of approach/method: How was the method changed/tailored to your situation?*

*Restructuring of organizational context: How were the work processes or organization structures changed when the method was taken into use?*

**Table 4.** Part 2 of the survey on DoI factors.

**2. Advantages of the method**

In case your paper describes a situation in which new methods (new to the method-user) were introduced, what aspects of relative advantage do you remember played a role in deciding to use the method you considered?

*Relative advantages: E.g., Social prestige, savings in time and effort, low initial cost, degree of economic profitability, increase in comfort/decrease in discomfort, immediacy of output (some may not be apparent prior to implementation), flexibility or capability of modification of an innovation,…*

In your case, were there any observability issues that played a role in method selection?

*Observability: How easy or hard it was to show that the method has a relative advantage to the previous situation, or to other methods considered?*

In your case, were there any issues related to compatibility of the method to needs, values and norms of the potential adopters and their social system, that played a role in method selection?

*Compatibility: What kind of advantage did the method have in relation to the needs, values in your team or organization?*

In your case were there any issues of trialibility or perceived complexity that played a role in method selection?

*Complexity: Was the new method easier to understand and use than the other methods that were considered, and did it affect the decision to adopt it?*

*Trialibility: Was it possible to try out the method beforehand, and did this affect the decision to adopt it?*

**Table 5.** Parts 3 and 4 of the survey on DoI factors

**3. People adopting a new method**

Attributes of adopters usually relate to organisational structure, - culture and – strategy. Can you remember (or have you written about) any of such factors that played a particular positive or negative role in your decision to adopt the described method or in not adopting an alternative method?

*Organisational structure, culture and strategy: E.g. size of organizations, aggressiveness and innovativeness of managers, level of information adopters have about existence and characteristics of an innovation, resistance to change, technical skills of users, competition (wanting to gain an advantage over competition), availability of resources in organization*

**4. Social system**

Can you remember (or have you written about) any factors such as related to the social system in which the method and adopters are embedded that played a particular positive or negative role in your decision to adopt the described method?

*Social system: E.g. social structure, norms, opinion leaders (e.g., lead method users that are seen as examples) and change agents (people promoting the use of a new method)*

*Communication channels: Was the method widely communicated in the outside world, or was it hard to find?*

Surveys were sent to all authors of 16 of the 18 papers. The two other papers were papers describing theoretical frameworks rather than cases of transfer (one of those was the paper presenting the DoI framework (Vermeeren & Cockton, 2013). Surveys

were filled in about the cases described in 10 out of the 16 papers. Authors of five papers had not responded, and one author indicated rightfully that the paper was not the type of paper that would match the survey questions. A total of 12 people filled in the survey for the 10 papers. In one case two authors collaboratively filled in the survey, in the other case an author had suggested to also contact another person who had not been involved in writing the paper, but had been involved in the described case. They each filled in the survey but they did so about different projects described in the same paper. In one case the respondent filled in the survey for each of the three projects described in the paper.

*Results of the survey*

The forms as they were filled in by the respondents can be found in Appendix E. The answers to the questions were analyzed by the WG4 coordinator, who interpreted and categorized the answers to the questions partly in a data-driven way with categories emerging during the analysis, and partly by grouping answers according to resource functions defined by WG1. The categorised results are presented below.

**A. Adopting a new method/approach**

Tables 6 through 10 present the answers relating to the process of adopting a new method/approach.

**Table 6.** Categories of answers about prior awareness of a method/approach; Legend for numbers in brackets: codes such as P1-2, refer to the second project described in paper 1.

How did awareness of existence of the method/approach affect considering it in your team or organization?

- No awareness of the method before project (7)
  - Team had no awareness; external in team made team aware (P1-1,P1-2, P8-1, P10)
  - Team had no awareness; internal in team was aware and made team aware (P12)
  - Team had no awareness; but actively searched for new approaches (P14, P16)
- There was awareness, but not of the specifics of the method (2)
  - Knowledge and skills of details brought in by external in team (P1-3)
  - Contact was sought with someone knowledgeable (P8-2)
- Awareness was present in at least a number of team members (P2, P7)
- There was awareness of the methods, but not of their importance (P15)
- Unclear whether there was any awareness before the project (P5)

In 7 out of 13 responses, the team had not been aware of a specific approach before the described project. In 5 out of the 7 cases they became aware because someone made them aware (4 external, 1 internal), in the other two cases they became aware by actively searching for methods. In 1 case there was awareness of existence of the method, but not of its importance, in 2 cases there was awareness but not enough knowledge to be able to use it. In one of those cases the knowledge was brought in by an external, in the other case by an internal. In 2 cases there was awareness, and in 1 case it was unclear.

**Table 7.** Categories of answers about Attitude forming towards a method/approach; Legend for numbers in brackets: codes such as P1-2, refer to the second project described in paper 1.

How did attitudes in (or outside) your team or organization affect considering the method/approach for use in your organization?

- Professionals (13):
  - Active, open, welcoming, curious, supporting (P2, P7, P8-1, P8-2, P14, P16)
  - Negative attitude (4)
    - Because of pride ("we know best") (P1-1)
    - Because of policy (only one contact person for customer) (P1-2)
    - Passive, ignoring (no pressure to do so, not seeing the need?) (P5)
    - Because of lack of confidence in method (P14)
  - Reluctance (1)
    - Because of image towards customer (not confront with unfinished materials) (P1-3)
  - Neutral (P10, P12, P15)
- Management (5)
  - Higher management was active, positive (P1-1, P1-2, P1-3, P5, P7)
  - Lower management showed resistance (P7)
- In-house users (1):
  - Positive (aware of problems and of need and benefit of change) (P5)

In 6 out of 13 cases there had been some negative attitude towards specific methods or approaches. This negative attitude is mostly (5 cases) at the level of professionals (method users), and in 1 case at the level of the management. As the papers all describe cases where new methods were used, the negativity was always overcome somehow. In 4 cases of a negative attitude amongst professionals this was overcome by the higher management actively pushing for change, in 1 case through support of in-house product users. In 1 case it was overcome by the active, curious and open attitude of other professionals in the team.

In the 1 case where there was negativity on the part of lower management this was overcome by higher management actively pushing for change, complemented by an active, open and welcoming attitude by professionals.

In 9 cases professionals were positive or neutral towards new methods, in 6 of those even active and supporting.

**Table 8.** Categories of answers about deciding on using a new method. Categorization partly based on WG1's resource functions; Legend for numbers in brackets: codes such as P1-2, refer to the second project described in paper 1.

Method characteristics matching the work: What aspects of the method were the key influencers in taking the decision of using (or not using) it?

- Axiological resource functions (3):
  - Objective data are valued (P1-2)
  - Comparability to results from other cases is important (P7)
  - Cost-effectiveness is important (P5)
- Scoping resource functions, method chosen because of (6):
  - Focus on user self-efficacy and enjoyment (P1-1)
  - Focus on learnablility (P1-1)
  - Focus on effectiveness (P1-3)
  - Focus on values (P2, P8-1, P8-2)
  - There was a wish to explore the scoping of the method (P12)
- Harvesting resource functions, method chosen because (4):
  - Can be done remotely (P1-1)
  - Provides objective data (P1-2)
  - It is non-user based (P5)
  - Provides qualitative data (P5)
  - Using internal expert evaluators makes it comparable to results from other cases they had done (P7)
- Directive resource functions, method chosen because (1):
  - Provides focus for the work (P2)
- Performative resource functions, method chosen because (1):
  - Convincing for executive decision makers (P5)
- Knowledge resource functions, method preferred because (1):
  - Builds upon available knowledge (P5)
- Integrative resource functions, method preferred because (1):
  - Easy recording of progress (P5)

In this question respondents were asked about key influencers in making a decision about a method. From the results it is clear that key influencers can be issues related to all kind of resource functions from values (axiological), to what is within the scope of a method (scoping), to what measures are used or to how data are collected (harvesting) etc. Based on these answers the assumptions about relating steps in the innovation decision processes to resource functions (as expressed in step 1c) seem too simplistic.

**Table 9.** Categories of answers about redefining a method when implementing a new method. Categorization partly based on WG1's resource functions; Legend for numbers in brackets: codes such as P1-2, refer to the second project described in paper 1.

When implementing the method, how was the method re-defined/changed/tailored to your situation?

- Approaches/methods were adapted (11):
  - Adapted for reasons related to axiological resource functions (2)
    - Priority given to project quality over speed while doing a lean start-up project (P2)
    - Measurement scale (-2 to +2) was added to Heuristic Evaluation (P7)
  - Harvesting resources adapted (7)
    - Elements of a method were tailored to the specific situation (4):
      - Additional heuristics were formulated in Heuristic Evaluation (P1-1, P7)
      - Users were asked what they would ant to do with a product and were then asked to do that, in Usability Testing (P1-3)
      - Cards in adjective card selection method were tailored to situation (P14)
      - Self-expression template were tailored to the situation (P14)
    - Questions were added to a question-driven method (1):
      - in Cognitive Walkthrough (P1-1)
      - in System Walkthrough (P5)
    - Changes were made to measurements (2):
      - Scores for impact analysis agreed during team evaluation (P5)
      - Measurement scale (-2 to +2) was added to Heuristic Evaluation (P7)
    - Changes were made in data collection strategy (2):
      - Team interview instead of individual interview (P10)
      - Algorithmic walking instead of random walking as data collection (P16)
  - Expressive resources (3):
    - Elements were added to the method (3):
      - in Worth Maps (P8-1, P8-2)
      - in Semiotic Engineering approach (P12)
- Elements of different methods/approaches were combined (1):
    - Contextual Inquiry used for identifying usability problems (P1-2)
    - Logs analysis based on task flow in personas (P1-2)
    - Asking users to make screen capture as part of remote survey (P1-2)

The results show that if methods/approaches are redefined, this is usually done at in their harvesting resource functions (7 out of 11 cases). In most cases this seems to have been done for tailoring the method to the specific situation at hand, however in two cases it seems to have been done for reasons that relate to axiological resource functions (e.g., making heuristics evaluation quantitative, giving priority to quality instead of speed while doing a lean start-up approach). In some cases one could also argue that methods were changed for reasons related to scoping resource functions. One could question whether there is a sharp distinction between tailoring as a standard aspect of using a method to the situation at hand a method to a specific situation (e.g., by necessarily tailoring cards in an adjective card selection method – tailoring then is an essential part of the method) and changing a method for use with a

specific type of products (e.g., scoping: adding heuristics focusing on mobile or on services). Combining elements from different methods can also occur, but was reported in one case only.

**Table 10.** Categories of answers about making changes in the organization due to implementing a new method; Legend for numbers in brackets: codes such as P1-2, refer to the second project described in paper 1.

How were the work processes or organization structures changed when the method was taken into use?

- Changes in the organization of teams (and not just for this project) (4)
  - Teams became smaller (because of agile and UX input from extern) (P1-3)
  - Local teams started to take the role of system re-designers (P5)
  - New autonomic and independent team was formed for expert evaluations (P7)
  - Experts in various departments started to become recruited as experts in HEs (P7)
  - UE first outsourced to student, then competence expanded through collaboration, then specialist was recruited for development projects (P15)
- Changes in responsibilities in the project team (specifically for this project) (1)
  - Everyone became responsible for a specific part of the data collection (P16)
- Changes in types of activities people were involved in (2)
  - Probably due to framework people were more out of the building, and there was different decision making and task prioritization (P2)
  - Usability evaluation gradually turned into service design (P5)

The results show that changes in the organization because of new approaches being introduced are no exception. In this sample of projects it occurred 6 times, and in 4 of those cases the change was not limited to the project about which was reported, but led to a more permanent organisational change. It is also remarkable that in a number of cases the side-effect was that people would start doing different activities in their current roles.

**B. Advantages of a method**

Tables 11 through 15 present the summarised answers to questions relating to advantages methods may or may not have: relative advantages, compatibility, observability, complexity and trialability.

**Table 11.** Categories of answers about relative advantages of methods/approaches. Legend for numbers in brackets: codes such as P1-2, refer to the second project described in paper 1.

Relative advantages: E.g., Social prestige, savings in time and effort, low initial cost, degree of economic profitability, increase in comfort/decrease in discomfort, immediacy of output (some may not be apparent prior to implementation), flexibility or capability of modification of an innovation,…

- Axiological resource functions (7 projects):
  - Costs related (4 projects):
    - Reduced time and effort (P1-1, P1-2, P2, P14)
    - Low initial cost (P1-1, P1-2)
  - Type of appreciated data related (3 projects):
    - Objectivity of output (P1-1)
    - Deeper information about UX (P14)
    - Provides systematic numbers on success and failures (P15)
  - Human relationships related (2 projects):
    - Enables direct interaction with users in person (P1-2, P1-3)
    - Fosters customer relations (P1-3)
  - Business related (1 project):
    - Confidentiality (as reason of using experts rather than users) (P7)
- Scoping resource functions (8 projects):
  - Gaining more (domain) insight (4 projects):
    - Insight into how software is integrated in user's work, work patterns and environment (P1-2, P5)
    - Deeper information about UX (P14)
    - Provides easy way of seeing patterns in data analysis (P16)
  - Provides information on specific qualities (1 project):
    - Provides satisfaction information (P1-3)
  - Method use was interesting as a scientific endeavour:
    - Addresses an intellectual challenge to find out about method (P8-1, P8-2)
    - Interesting case study for publication (P12)
- Harvesting resource functions (7 projects):
  - Data characteristics (7 projects):
    - Provides objective data, systematic numbers (P1-1, P15)
    - Provides graphical information (P1-2)
    - Generates ideas and solutions (P5, P14)
    - Provides structure in data (P16 )
    - Provides easy mapping of city centre, through geotagged materials (P16)
    - Provides precise and actionable feedback (P7)
    - Data can be used for online data analysis service (P16)
  - People related (3 projects):
    - Manageable to cover wide number of users (P1-2, P5)
- Directive resource functions (7 projects):
  - Time-related (2 projects):
    - Immediacy of output, data immediately available in process (P1-1)
    - Actionable output, data in a form that it can readily be used in process (P7)

- Process structure related (5 projects):
  - Flexibility (P1-2, P2)
  - Before evaluation it already generates ideas and solutions (P5)
  - It formalizes evaluations (P15)
  - Provides structure for researchers when collecting data (P16)
- Performative resource functions (1 project):
  - Letting customers experience new functionality, while they continue their work. Gives head start when they start using the new tool (P1-3)
- Knowledge resources (2 projects):
  - Helps company to understand UX as part of service quality (P5)
  - Develops internal competence (P7).

When asked for key influencers in deciding on adoption of methods, axiological, scoping and harvesting resource functions were mentioned most, although other functions were also mentioned. The question above, about relative advantages of methods shows a similar range of reasons: axiological, scoping and harvesting with directive advantages mentioned frequently as well.

**Table 12.** Categories of answers about compatibility of methods/approaches with the organization. Legend for numbers in brackets: codes such as P1-2, refer to the second project described in paper 1.

Compatibility: What kind of advantage did the method have in relation to the needs, values in your team or organization?

- Axiological resource function (3 projects):
  - Trust and confidence of existing users important (P1-3)
  - Team valued trying innovation of methods (P12)
  - High quality, validity of results was crucial (P15)
- Scoping resources function (3 projects):
  - There was an understanding of a need to better understand users, and that fits the approach (P1-2)
  - Focus on value was appreciated (P8-1)
  - Focus on UX rather than usability (P14)
- Harvesting resource function (1 project):
  - There was familiarity with elements of the methods (P16)
- Emotional resource function (1 project):
  - Framework served as mission and vision, fitted mindset (P2)

When asked about what was important for compatibility of the method/approach with the organization, answers mainly focused on axiological (3 out of 8 projects) and scoping (3 out of 8 projects) resource functions. In itself this would be expected. However, again it is striking that when asked for the key influencers in deciding for a method, the focus was relatively more on scoping and harvesting resource functions.

**Table 13.** Categories of answers about observability of the effects of using a method/approach. Legend for numbers in brackets: codes such as P1-2, refer to the second project described in paper 1.

Observability: How easy or hard was it to show that the method has a relative advantage to the previous situation, or to other methods considered?

- Positive about observability of method (6 projects):
  - Advantages of method a priori observable, because (3 projects):
    - Results were shown to be more fine-grained than without method (P12)
    - Method provided new type of data, that previously was not gathered (P15)
    - Common-sense tells that the adaptation of the method made it more systematic (P16)
  - Project results became more convincing because (3 projects):
    - Videos are convincing and easy to understand (P1-3)
    - Cross-validation of various methods (UT and AttrakDiff) made results convincing (P2)
    - Attending a method session made it convincing for manager (P10)
- Negative about observability of method (3 projects):
  - Advantages of a method a priori not observable, because (2 projects):
    - Difficult to see beforehand because Worth Maps are so abstract (P8-2)
    - Cost-benefits were hard to determine, while costs were relatively high (P15)
  - Problems in making project results convincing, because (1 project):
    - Results are very subjective (individual) and qualitative, so difficult to show advantages to others (P14)

When asked about observability of effects of using a method two types of answers were given: 1) answers relating to determining whether the type of results itself makes clear what the advantage of a method is, and 2) answers relating how the results from a particular project become more convincing because the method offered others involved in the project a way to get insight into the quality of the data. For this second type of observability trialibility of a method is important for increasing the chance of adoption, as it is about observability of results of a particular project. For the first type of observability, trialibility is not necessary, as this is about the inherent characteristics of the type of data not about the particular results of a project.

**Table 14.** Categories of answers about triability of a method/approach. Legend for numbers in brackets: codes such as P1-2, refer to the second project described in paper 1.

Trialibility: Was it possible to try out the method beforehand, and did this affect the decision to adopt it?

- Was not an issue because (4 projects):
  - The method was new, the project as a whole was about trying out the method (P2)
  - Variant of the method was already used, and the changes did not require substantial additional investments in time or effort and did not lead to much risk (P7, P10)
  - We invented the method (P16)
- The reported study was the trial, by fully using it (P8-1, 8-2)
- One variant was tried out first, before the more complex second variant was tried (P12)
- Pilot tests were conducted (P14)
- First tried without investing in extra personnel, by hiring a student (P15)

In a few cases the project itself focused on trying out a new method, or the method was invented by the method users. In other cases variants of existing methods were used. One respondent interpreted pilot studies as trials of a method. This could be the case if the pilot study is the first use of a method. Difference between pilot study and trialability as intended here, however, is that usually in case of a pilot study one has already decided to use a method, and wants to find out whether it still needs refinement or tailoring.

**Table 15.** Categories of answers about perceived complexity of a method/approach. Legend for numbers in brackets: codes such as P1-2, refer to the second project described in paper 1.

Complexity: Was the new method easier to understand and use than the other methods that were considered, and did it affect the decision to adopt it?

- There was complexity, but (6 projects):
  - Complexity was handled by someone originally external to the team (5 projects):
    - Primary researcher developed own experience guided by academic experts (P1-1, P1-2, P1-3)
    - The inventor was present, and this was needed (P8-1)
    - People were first hired in for that, and later competence was expanded (P15)
  - That was the reason why the method was explored (P8-2)
- Aspects of complexity mentioned (5 projects):
  - Approach was easy to communicate, understand and tailor, but not the framework, which was abstract (P2)
  - Methods were easy to follow (P5, P16)
  - Results were easy to understand (P10)
  - Terminology was a hindrance to adoption by other team members (P12)

All reported cases in the survey were about projects where new methods, or adaptations of methods had been used. Therefore, in the survey there are no answers relating to perceived complexity of methods hindering adoption. However, in the answers to 5 projects (3 people in this case) something was said about how measures

had been taken to overcome problems of complexity, i.e., by getting the expertise from people external to the team. The results also show that complexity can be in different elements of a method: in a framework on which a method is based, in the procedure ('method was easy to follow'), in understanding the results or in its terminology.

### C. People adopting a new method

Table 16 presents results related to attributes of people (or organizations) adopting a new method.

**Table 16.** Categories of answers about attributes of people (or organizations) adopting a method/approach. Legend for numbers in brackets: codes such as P1-2, refer to the second project described in paper 1.

Can you remember (or have you written about) any factors relating to organisational structure, - culture and – strategy that played a particular positive or negative role in your decision to adopt the described method or in not adopting an alternative method?

- Characteristics of managers (6 projects):
  - Actively supporting managers helped adoption (P1-1, P1-2, P1-3, P7, P8-1)
  - Managers had not expected change of method, but just a report that would lead to improvements (P5)
- Characteristics of the larger organization (7 projects):
  - Small organization with people that trust each other, which helped (P10)
  - Growth of the organization led to managers not trusting team's gut feeling, but trusting numbers to base decisions on (P14)
  - Change had not happened before because in-house users were located remotely, and needs were easier to ignore by other department(P5)
  - Availability of resources in organisation helped (P8-1, P15)
  - Agile software development environment, led to very short iterations and absence of possibility for long-term studies (P14)
  - Local in-house users showed positive attitude to cooperate (P5)
- Team characteristics (10 characteristics):
  - Culture of no need for change, hindered adoption (P1-1, P1-2, P1-3, P5, P7)
  - Culture of innovativeness and open-mindedness in the team (P2, P12)
  - Skills of method users was important (P8-1, P16)
  - Multi-disciplinarity of teams makes method more fruitful (P8-2)
- Pressure from market position/competition helped (P1-1, P1-2, P1-3, P2, P5, P8-1)

As was mentioned in the answers to questions about attitude, the fact that management actively supports introduction of new methods helps in introducing the methods. Market positions and wanting to gain competitive advantage also helps in introducing new methods; they are probably one of the reasons why managers are supportive. Team characteristics obviously can also play an important role in adoption of methods. This can relate to culture, composition and/or available skills in the team.

Characteristics of the larger organization around a team may also influence introducing new methods (e.g., size of the organization and whether decisions are taken based on numbers, gut feelings or trust). Availability of resources for new methods is another factors that plays a role.

**D. Social system**

Table 17 presents answers to questions related to the social system in which a method or adopter was embedded.

**Table 17.** Categories of answers about social system in which method or adopter were embedded. Legend for numbers in brackets: codes such as P1-2, refer to the second project described in paper 1.

Can you remember (or have you written about) any factors relating to the social system in which the method or adopters were embedded that played a particular positive or negative role in your decision to adopt the described method or in not adopting an alternative method? E.g. social structure, norms, opinion leaders (e.g., lead method users that are seen as examples) and change agents (people promoting the use of a new method)

- Higher management support (3 projects):
  - Higher management supported when some stakeholders opposed to the new method (P7)
  - CEO trusted researcher and staff trusted CEO (P10)
  - Boss had been familiar with method in the past, this helped in getting permission (P12)
- Decision to bring in externals for the new method (5 projects):
  - Researcher was brought in by company as change agent, realizing lack of expertise (P1-1, P1-2, P1-3, P8)
  - Students were first brought in for expertise, then collaboration with industry, then recruit new person (P15)
- Fact that it was collaboration between academics and industry influenced choice of methods (P1-1, 1-2, 1-3)
- There already was a culture of in-house users already having an unofficial support system and inventory of workarounds (P5)

Again, the answers show that higher management can play an important role in introducing new methods. Often indirectly, by supporting the need for change or by trusting and supporting change agents. Bringing in new knowledge or expertise for using a new method may occur by bringing in externals (as was also noticed in the answers about how introducing a method changed the organization).

Questions about the influence of communication channels in relation to method use did not lead to insightful answers. Respondents mainly mentioned how in general the methods are communicated, but had written nothing about whether this had an effect on their adoption process.

### 2.3 Discussion

The various steps WG4 had taken gradually converged to a proposed framework, which was then empirically studied by sending out a survey to people who had submitted real-life cases of transfer. Responses to the framework provided a detailed insight into which real-life factors play a role in method adoption, as well as in who are involved and how. Two issues should be mentioned in relation to the validity and generalizability of the results. Categorization of responses to an open-ended questionnaire inherently is a subjective endeavour. Although in many cases well-defined categories from published literature of WG1 were used (e.g., Vermeeren & Cockton, 2013), one cannot exclude that if someone else would do the categorisation again, they would end up with different categories, or a different categorisation of items. On the positive side, one could argue that this would only lead to additional insights into the situation, rather than to invalidating the insights gained thus far. An additional issue is the selection of real-life cases. Obviously this is a skewed sample, in the sense that it only includes cases from people who go to conferences like CHI, and in that it only concerns stories about adoption of methods, or at least trials of methods, but not about rejection of methods. For a more complete view on transfer this would also be necessary. Again, we suppose that collecting such additional data would not invalidate the current findings, but would make the framework presented in the next section more comprehensive.

## 3 Integrated D&E framework

An integrative framework is proposed for diffusion of methods across contexts (Figure 16). The DIDEMI (DIffusion of Design and Evaluation Methods Innovations) framework depicts the adoption of new methods as a process from agenda setting, via decision making to implementation, after which a method becomes part of the normal practices of an organization. Next to the process arrows, clouds depict the influential factors providing input to the process and affecting it. These include: adopter attributes as well as method resources and how these are perceived by the adopter organization.

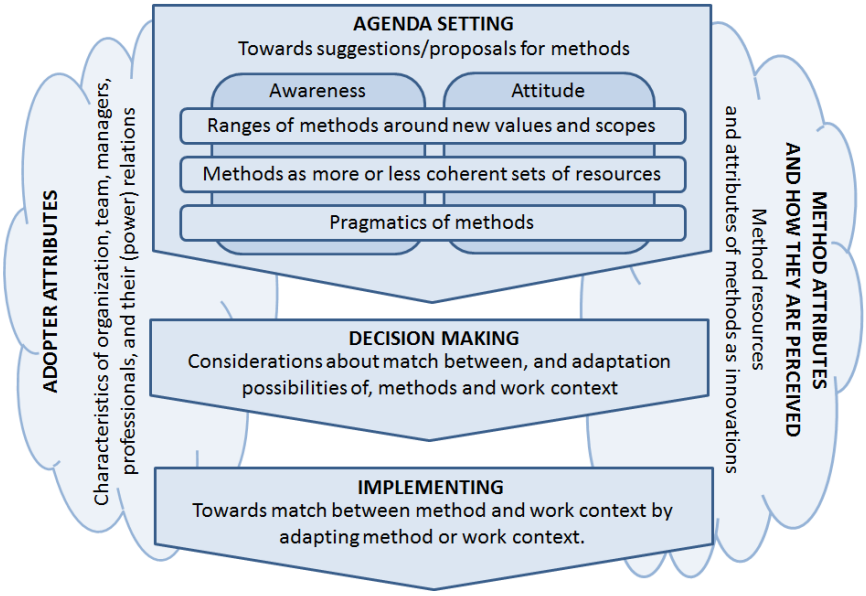



**Fig. 16.** Framework for DIffusion of Design &Evaluation Method Innovation (DIDEMI)

### 3.1 Description of the framework

The adoption process starts with the *agenda setting stage*. In this stage the idea for 'something new' arises. In line with Rogers (1983) we suggest that the idea for a new method can arise either from a recognised need or from coming across a new method or new range of methods. The three horizontal bars in the *agenda setting* arrow depict various levels at which this can occur. For example, from the survey it has become clear that in a number of cases management had recognised that something needed to be changed, in the way of working in an organization. This can, for example, be inspired by analysing an organization's market position and by concluding that the client (or user) needs to get a more prominent place in the work. In other words the organization's values change (e.g., the product's effectiveness-in-use is no longer a sufficient condition for success, and usability and user experience have become crucial as well). This then inspires a search for a new *range of methods* (Figure 16, upper bar in the agenda setting arrow) that address these values (axiological resource functions), and which address new issues such as designing for experience (scoping resource functions). In terms of the DIDEMI framework, the manager has become *aware* of a new need, and has a positive *attitude* towards ranges of methods that address this need. In the same organization, professionals may not always be *aware* of that same need. Moreover, they may simply hate having to go through a change again (negative *attitude* towards change), or have a negative attitude towards this new *range of methods* ("we know best how to do our work, we never hear customers complain!"). An external brought into the team by managers as change agents will obviously have a positive *attitude* towards the new range of methods (that is what the

person is hired for), and will have the task to propose specific new *methods* (Figure 16, middle bar). In doing so they may meet negative attitudes of professionals at the level of *methods* ("we don't want to confront our customer in user tests with products that are not finished") or of some of the *pragmatics of the method* (Figure 16, lower bar) proposed ("this takes too much time"). Changes agents may also have their specific *attitudes* towards methods, and may possibly not be aware of all the *pragmatics* of some of the candidate methods. Summarizing, awareness of, and attitudes towards needs for change and ideas/suggestions for change can occur at various level of specificity as depicted by the three horizontal bars in the agenda setting arrow. The cloud depicted on the left of the figure depicts the influence adopter factors such as the organization, the team, (power) relations may have on the process. This can occur at various stages in the process.

Once one or more methods have become candidates to seriously consider, the *decision making stage* starts. Rogers (1983) defines this (matching) stage as the stage "in which an organization is trying to figure out whether it seems worthwhile to adopt the innovation or not and tries to imagine the consequences of the innovation when implemented in the organization". In the answers to the survey this is reflected in the answers to the question about key influencers for adopting a method, as well as in the answers to the questions on relative advantages of a method. Another related question concerns the compatibility of a method with team and organizations values and needs. In response to the compatibility question, responses focused on axiological and scoping resource functions, which is understandable as this is a question of how a method relates to 'what an organisation is or stands for'. Presumably such axiological and scoping issues are the first considerations a manager would give to the proposal of a new method, early on in the adoption process. When asked for the key influencers, respondents also indicate they focus on axiological and scoping resource functions, but harvesting resource functions are added to that. On closer look this mainly focuses on general issues of harvesting, such as 'can it be done remotely?', 'does it provide objective data', etc. Presumably these are still considerations from early on in the decision making process. When asked for relative advantages of methods, respondents add another substantial category to the other three main categories, namely directive resource functions. It is clear that at this point all details of how a method fits into the work processes are considered. It seems that if a method has to be selected in an organization, agenda setting and the initial stages of the decision making stage have a tendency to focus on axiological, scoping and harvesting resource functions, with a likelihood of managers to be involved initially. Additionally it seems that when there is more certainty on where the decision process may be heading, directive resource functions also will start being considered. It must be stressed, however, that based on the survey results, it is also clear that method selection not always follows this filtering process, but can also just start by someone coming across an interesting new method that seems to focus on the right values and has the right scoping, and which is then just tried out to explore it's worth. *Observability* and *trialibility* of new methods may also play a role in selecting a method, but in the answers we received to the questionnaire, these factors didn't come across very strongly. *Perceived complexity* of a method, proved to be important, but

in the reported cases this often led to organizations hiring people to get the knowledge needed for using a method; it never led to not adopting a method. The resource functions, as well as how they are perceived by the adopters are graphically indicated in the figure by the cloud on the right.

Once a decision of adopting a method has been made, the *implementation* stage starts. In this stage the method has to be made to work, which usually requires adaptation and tailoring of a method, often in combination with changes in people's activities or even in composition of teams, and/or dealing with customers (see responses to questions on changes in processes or organization). Changes in methods at this stage usually are at the level of harvesting resource functions, but at a more detailed level than in the decision making stage. In the decision making stage this concerned general issues of harvesting such a 'can it be done remotely?', at this stage it concerns the details of how to do the harvesting (adding heuristics to a standard list of heuristics, deciding to interview multiple people of a team at once, instead of individually, etc.

### 3.2 Positioning of MCDMs and results of the other WGs

As has become clear in the previous chapter the results of the other WGs played an important role in composing this framework, it is based on their work. WG1's resource functions played a major role as will be clear by now. They find their place in the cloud at the right hand side of the framework. WG3's quality-in-use attributes are also in that cloud. In terms of WG1's resource functions these can be interpreted as a specific set of axiological and scoping resource functions. WG2's focus on the transfer process rather than on a sole matching of attributes is obviously reflected in the central process (depicted as the arrows). As had become clear in WG4's step 1c the process flow WG2 proposed seems to be specifically focusing on certain types of organizations, making it unsuitable to include it in the DIDEMI framework as such. Nevertheless for the situations it has been designed for, it can be very useful. Therefore, WG4 chose to present a more abstract and general process, and sees WG2's process model as a specific instantiation of that more general process.

The MCDMs potentially find their place at all stages in the process, where method attributes are considered in relation to factors of the work context.

## 4 Impacts achieved and/or Projected

*Published*:

- Law E. L.-C., Hvannberg E., Vermeeren A.P.O.S., Cockton G., Jokela T. Made for Sharing: HCI Stories of Transfer, Triumph & Tragedy. Ext Abstract CHI'13. ACMPress (2013).CHI workshop
- Vermeeren , A.P.O.S., Cockton, G. Facilitating the take-up of new HCI practices: a 'diffusion of innovations' perspective. Workshop W20: Made for Sharing: HCI Stories of Transfer, Triumph and Tragedy @CHI 2013; 04/2013

*Planned*:
Special Issue Romanian Journal of HCI
Article presenting the framework in the special issue.

## 5 Conclusion and Future Work

A framework was developed for providing insight into the factors that play a role in the selection of new methods in organizational settings in various contexts. It was developed in a series of workshops with experts, and was finalized based on a set of published case studies of new method adoption in real-life, and on a survey which was sent to people involved in those cases. Part of WG4's work was also to study method selection tools and MCDMs. Application of MCDM was explored empiricially with simulated data. It was decided that method selection tools, especially with embedded MCDMs would not be feasible for development within a COST projects. Moreover there were doubts whether substantial use of MCDMs for large sets of attributes would be pragmatically feasible at all.

Future work based on WG4's work could consist of optimising the framework by studying additional real-life cases, such as those in which methods were rejected for adoption, and cases from professionals who are not interested in publishing at venues like CHI conferences. After optimisation of the framework, it could be explored what the consequences of the framework are for a method selection tool to be developed.

## 6 References


1. Askarany, D. (2003) An Overview of the Diffusion of Advanced Techniques. In B. F. Tan (Ed.), *Advanced Topics in Global Information* (*Vol. 2*), pp. 225-250). IDEA Group Publishing.
2. Brans, J. P., Vincke, P., and Mareschal, B. (1986). How to select and how to rank projects: The Promethee method. *European Journal of Operational Research*, 24(2), 228–238.
3. Card, S., Moran, T. P. and Newell, A. (1986). *The Psychology of Human Computer Interaction*. Lawrence Erlbaum Associates.
4. Furniss, D. (2008). *Beyond Problem Identification: Valuing methods in a 'system of usability practice*. PhD Thesis. UCL. Available in http://dl.dropbox.com/u/2191887/My Publications/2008_Furniss_PhDThesis.pdf
5. J. Nielsen and D. Sano, "SunWeb: user interface design for Sun Microsystem's internal Web," Computer Networks and ISDN Systems, vol. 28, no. 1–2, pp. 179–188, Dec. 1995.
6. Kujala, S., Roto, V.., Väänänen-Vainio-Mattila, K. and Sinnelä, A. (2011). Identifying hedonic factors in long-term user experience. In *Proc. DPPI '11*
7. Lahdelma, R., and Salminen, P. (2001). SMAA-2: Stochastic Multicriteria Acceptability Analysis for Group Decision Making, *Operations Research*, 49(3), 444–454.
8. Law E. L.-C., Hvannberg E., Vermeeren A.P.O.S., Cockton G., Jokela T. (2013). Made for Sharing: HCI Stories of Transfer, Triumph & Tragedy. In *Proc. Ext Abstract CHI'13*.
9. Mareschal, B., and De Smet, Y. (2009). Visual PROMETHEE: Developments of the PROMETHEE & GAIA multicriteria decision aid methods,. In *Proc. IEEE International*

*Conference on Industrial Engineering and Engineering Management* (Vol. 2), pp. 1646–1649.
10. Melo, P. and Jorge, L. (2013). Quantitative Support For UX Methods Identification: How Can Multiple Criteria Decision Making Help? *Universal Access in the Information Society* (UAIS), in print.
11. Nielsen, J. (1994). Usability inspection methods. In *Proc. CHI '94*, pp. 413–414.
12. Nielsen, J. and Molich, R. (1990). Heuristic evaluation of user interfaces. In *Proc. CHI'90*, pp. 249–256.
13. Norman, G. (2010). Likert scales, levels of measurement and the 'laws' of statistics. *Advances in Health Sciences Education: Theory and Practice*, 15(5), 625–32.
14. Rogers, E. M. (1983). *Diffusion of innovations*. The Free Press.
15. Siskos, Y. Grigoroudis, E., and Matsatsinis, N. (2005). UTA methods. In J. Figueira, S. Greco, and M. Ehrgott (Eds.), *Multiple Criteria Decision Analysis: State of the Art Surveys (*pp. 297–344). Springer.
16. Vermeeren , A.P.O.S., and Cockton, G. (2013). Facilitating the take-up of new HCI practices: a 'diffusion of innovations' perspective. In *Proc. CHI'13 Workshop W20: Made for Sharing: HCI Stories of Transfer, Triumph and Tragedy*
17. Wixon, D., Holtzblatt, K. and Knox, S. T. (1990). Contextual Design: An Emergent View of System Design," In *Proc. CHI'90*.
18. Yarbus, A. L. (1967). *Eye movements and vision*. Plenum Press.

# 7 Appendices

Appendix A. Bertinoro workshop (WS1) report
Appendix B. WS2 slides introducing the task at the London workshop
Appendix C. Materials for Coimbra workshop
Appendix D. Survey format
Appendix E. Responses to the survey.

European Science Foundation provides and manages the scientific and technical secretariat for COST

COST is supported by the EU RTD Framework Programme