A profile-aware methodological framework for collaborative multidimensional modeling

Published Version:
A profile-aware methodological framework for collaborative multidimensional modeling / Amir Sakka; Sandro Bimonte; Stefano Rizzi; Lucile Sautot; François Pinet; Michela Bertolotto; Aurélien Besnard; Noura Rouillier. - In: DATA & KNOWLEDGE ENGINEERING. - ISSN 0169-023X. - STAMPA. - 131-132:(2021), pp. 101875.1-101875.23. [10.1016/j.datak.2021.101875]

Availability:
This version is available at: https://hdl.handle.net/11585/812718 since: 2022-01-31

Published:
DOI: http://doi.org/10.1016/j.datak.2021.101875

Sakka, A., Bimonte, S., Rizzi, S., Sautot, L., Pinet, F., Bertolotto, M., . . . Rouillier, N. (2021). A profile-aware methodological framework for collaborative multidimensional modeling. Data and Knowledge Engineering, 131-132

The final published version is available online at: **https://doi.org/10.1016/j.datak.2021.101875**

A profile-aware methodological framework for collaborative multidimensional modeling

Amir Sakka[a], Sandro Bimonte[b], Stefano Rizzi[c], Lucile Sautot[d], François Pinet[b], Michela Bertolotto[e], Aurélien Besnard[f], Noura Rouillier[g]

[a] *IRIT, Université Capitole, Toulouse, France*
[b] *University of Clermont Auvergne, INRAE, TSCF, Clermont Ferrand, France*
[c] *DISI, University of Bologna, Bologna, Italy*
[d] *UMR TETIS, AgroParisTech, Montpellier, France*
[e] *University College Dublin, Belfield, Dublin, Ireland*
[f] *LPO Aquitaine, Villenave-d'Ornon, France*
[g] *CESCO, Muséum National d'Histoire Naturelle, Paris, France*

Abstract

Multidimensional modeling, i.e., the design of cube schemata, has a key role in data warehouse (DW) projects, in self-service business intelligence, and in general to let users analyze data via the OLAP paradigm. Though an effective involvement of users in multidimensional modeling is crucial in these projects, not much has been said about how to establish a fruitful collaboration in projects involving numerous users with different skills, reputations, and degrees of authority. This issue is especially relevant in citizen science projects, where several volunteers can contribute their requirements despite not being formally-trained experts in the application domain. To fill this gap, we propose a framework for collaborative multidimensional modeling that can adapt itself to the profiles and skills of the actors involved. We first classify users depending on their authoritativeness, skills, and engagement in the project. Then, following this classification, we identify four possible methodological scenarios and propose a profile-aware methodology supported by two sets of quality attributes. Finally, we describe a Group Decision Support System that implements our methodological framework and present some experiments carried out on a real case study.

Keywords: Data warehouse design, collaborative systems, quality dimensions

1. Introduction

Multidimensional modeling is the activity of designing multidimensional schemata for data in such a way that they can then be used by analysts to answer their queries of interest. Since the multidimensional model relies on the metaphor of *cubes* to describe phenomena of interest, the resulting schemata are also called cube schemata. Representing data in multidimensional form enables analysts to access them using the OLAP (On-Line Analytical Processing) paradigm, which has been fruitfully adopted for more than 20 years to support decision-making processes.

Multidimensional modeling is necessary in different design contexts. First of all, it is a key activity in the design of data warehouses (DWs), which enable advanced exploration and visualization of huge volumes of data via OLAP tools [1]. The approach adopted in DW systems is often called *schema-on-write*, because source data are transformed and stored in multidimensional form using schemata decided at design time. Multidimensional modeling is also necessary to enable OLAP analyses on schemaless databases in so-called *schema-on-read* approaches, where data are left in their original formats (e.g., graph-based or document-based) within a *data lake* until they are accessed by the user (typically, a data scientist) [2]. Similarly, multidimensional modeling is required in *self-service business intelligence*, where the search and analysis of data is accomplished by users on-the-fly without any mediation by analysts, designers, or programmers [3]. In schema-on-read approaches the multidimensional schema to be adopted for analyses is not decided at design time but at query time; still, multidimensional modeling is important to let users choose a useful perspective for analyzing data through OLAP tools. Importantly, though for simplicity in the rest of the paper we will always refer to DW projects as the context for multidimensional modeling, all the arguments can be smoothly extended to schema-on-read approaches and self-service business intelligence.

Two main groups of actors are typically involved in DW projects:

- *Users*, who access the DW to analyze data. Typically, they have good to excellent skills in the application domain, and may be more or less authoritative depending on their role. They may have some (basic or even advanced) knowledge of OLAP and multidimensional modeling.

- *Designers*, i.e., Information Technology (IT) people with skills in DW design in charge of managing the technical issues of the project.

Clearly, an effective involvement of users is one of the keys to the success of DW projects [4]. Some existing approaches to design DWs (e.g., [5]) have been shown to be effective to this end, with specific reference to projects taking place in companies, where few and authoritative users are typically involved.

This scenario is today made more complex by the spread of data acquisition systems (embedded sensors, smartphone cameras, social networks, etc.) and by the democratization of information technology, which are making citizen observatories and citizen science more and more common. *Citizen science* has been defined as an "online, distributed problem-solving and production model" [6]. It typically relies on *crowdsourcing*, i.e., on a large pool of people gathering inputs such as ideas, funding, etc. On the one hand, the role of crowdsourcing in citizen science projects consists of having volunteers provide data and collaboratively edit them, despite not being formally-trained experts in the application domain. On the other, it extends the borders of decision making and introduces another type of actors in DW projects: volunteers, who —differently from users in traditional projects— are often not authoritative, not fully engaged in the project, and have poor or no skills in multidimensional modeling.

Though the innovative collaborative model provided by citizen science promotes inclusion and awareness by making data more widely accessible for analyses, it also raises a methodological challenge. Indeed, coping with crowdsourcing in DW projects requires establishing a fruitful *collaboration* among a large number of contributors with different skills, reputations, and degrees of authority. So far, not much has been said in the literature about this. In an attempt to fill this gap, the authors of [7] investigated the collaborative design of cube schemata, proposing a methodology that relies on three groups of users: *volunteers* (non-authoritative users who express some preliminary analysis needs), *committers* (authoritative users who collaboratively validate the volunteers proposals), and *DW experts* (who take care of all technical issues related to the development of the DW). Volunteers can give their feedback about the cube schemata; to achieve a consensus about the elements of these schemata, they use a *Group Decision Support System* (GDSS), i.e., a software tool that allows the management of group meetings with mathematical methods for finding solutions to problems that are unstructured in nature [8].

However, some experiments carried out on a real case study have highlighted some limitations of such a methodology, mainly related to the difficulty in coping with the varying skills of users in the definition of analysis requirements [7, 9]. Some relevant issues that remain open are:

- "How to organise and prioritize the analysis needs of several users?"
- "How to deal with —and take advantage of— the specific skills and profiles of the users involved in the design process?"
- "How to cope with the broad geographic distribution of users and with the timeline of their interventions?"
- "How to deal with *vandalism* [10], i.e., with the participation of volunteers with malicious intents associated to religious, professional, and political goals?"

To answer these questions, in this paper we propose a framework for collaborative multidimensional modeling that can adapt itself to the profiles and skills of the actors involved. Our work provides the following original contributions:

1. We classify users based on their authoritativeness, distinguishing between (non-authoritative) *end-users*, who just express analysis requirements, and (authoritative) *decision-makers*, who validate the end-users requirements and integrate them with their own requirements. Then, we further classify decision-makers and end-users based on their skills and engagement in the project, respectively (Section 4).
2. We identify four possible methodological scenarios based on the above classification (Section 4).

3. Taking as a reference the most challenging methodological scenario, we propose a profile-aware methodology that extends the one in [7] by taking into account the varying skills of the users involved (Section 7). To this end we identify (i) a metric aimed at assessing how trustworthy each schema element is, also considering the reliability of volunteer end-users (Section 5), and (ii) a set of quality attributes[1] aimed at supporting decision-makers without multidimensional modeling skills in inspecting cube schemata and detecting errors (Section 6).

4. We describe a web-based GDSS, called GROUDA, which implements our methodological framework by tailoring it to the specific users' profiles (Section 8).

5. We discuss some experiments carried out in the context of a real case study in biodiversity to test our approach (Section 9).

The paper outline is completed by a discussion of the related literature (Section 2), a description of our working example in biodiversity (Section 3), and some concluding remarks (Section 10).

2. Related work

2.1. Multidimensional modeling

The approaches to multidimensional modeling were originally classified into two categories [12], data-driven and requirement-driven. *Data-driven* approaches design the schemata starting from a detailed analysis of the data source schemata; user requirements impact on design by allowing the designer to select which chunks of data are relevant for decision making and by determining their structuring according to the multidimensional model [13, 14]. *Requirement-driven* approaches start from determining the information requirements of end users, while how to map these requirements onto the available data sources is investigated only afterwords [15, 16]. A third category of *hybrid approaches* has been then emerging as a mixture of data- and requirement-driven ones; in this case, both data source schemata and user requirements are used at the same time [17, 18]. Finally, in *query-driven* approaches, a multidimensional schema is created starting from the set of OLAP queries the users are willing to formulate; these queries are specified using either SQL statements [19], MDX expressions [20], or query trees [21]. We refer the reader to [22] for a nice survey of the approaches to multidimensional modeling. Interestingly, although some approaches (e.g., [23, 24]) take the decision-makers' profile into account to personalise OLAP analyses, they do not cover design.

With specific reference to requirements elicitation, several works propose methods and tools to express the analysis requirements and translate them

[1] In the literature these are often referred to as *quality dimensions* [11]; we prefer not use the term "dimension" to avoid confusion with the dimensions of a cube.

into multidimensional schemata (see [25] for an up-to-date survey). Examples of tools adopted to this end are UML use case diagrams, natural language, and interviews [26, 27]. However, these tools appear ineffective when decision-makers have poor skills in DW and OLAP, since they are either too complex (e.g., UML or extensions of the Entity/Relationship model [28, 18]), or too ambiguous (e.g., natural language). For this reason, pivot tables have been proposed as easily understandable and non-ambiguous tools that can be used by any kind of decision-makers [25].

Collaborative techniques and tools have already been applied to software engineering, leading to the concept of *social tools for software engineering* [29, 30]. For instance, collaborative tools, such as Wiki and forums, are used in the context of programming activities for software development to achieve fast and easy content creation [31]. Wiki has also been used for DW testing activities [9].

To the best of our knowledge, only [7] investigates the collaborative design of cube schemata in crowdsourcing scenarios taking the profile of volunteers into account. In particular, the authors propose a DW design methodology where three groups of users have been identified: *volunteers*, *committers*, and *BI experts*. The main goal of this methodology is to collect the users' feedback about cube schemata and achieve a consensus about its elements using mathematical methods provided by GDSSs [8]. However, the approach in [7] has the drawback of not properly accounting for the variability of skills among all actors participating in the system design. In this paper we propose a methodology that addresses such a limitation.

2.2. Multidimensional modeling quality

Quality is a fundamental aspect of all software applications and information systems. Several works investigate the quality attributes of information systems —including DWs— from both the data and schema points of view [11], e.g., in terms of completeness, accuracy, etc. These attributes are often associated to quality metrics. A quality metric can be objective or subjective, and it represents a measurable aspect of a quality attribute. These measures are then used to assess the quality attribute according to an expected value [32].

According to [33], quality in DWs can be investigated from three points of view: *presentation quality*, *data quality*, and *schema quality*. The existing literature has mainly addressed data quality and schema quality. Here we will only focus on the approaches focused on schema quality, which are relevant to our work.

In [34] the authors define a framework for the analysis of the quality of multidimensional schemata according to three points of view: specification (the designer's), usage (the decision-maker's), and implementation (the developer's). In particular, from the specification point of view, the authors focus on *readability* and *expressiveness*, delivering the following metrics: *non-redundancy* (i.e., minimality), *factorisation degree* (reuse of a hierarchy for different facts), *zoom-in zoom-out facility* (related to the maximum depth of hierarchies), *fact richness* (related to the number of measures), *fact dimensioning* (related to the number of dimensions), *fact analysability* (overall number of levels), *fact summarizability* (number of aggregation functions applicable), *simplicity* (related to the total number of concepts expressed in a schema), and *correctness* (related to the number of errors found in a schema). Note that all these metrics evaluate a cube schema in its entirety, thus they do not provide indications on each single multidimensional element.

The schema quality metrics proposed and validated in [33] are related to the *understandability* of multdimensional schemata, which is strictly connected to the readability mentioned above. Examples of these metrics are the number of dimensions of a fact, its number of measures, and the maximum depth of fact hierarchies. These metrics are evaluated with reference to either single multidimensional elements, single multidimensional schemata, or set of multidimensional schemata forming a DW [35]. Further metrics concern the *maintainability* of multidimensional schemata and evaluate the quality of a multidimensional schema with respect to its ability to sustain changes during an evolution process. Not surprisingly, empirical evidences based on simulated sequences of evolution events suggest that the vulnerability of a fact or dimension to future changes mostly depends on its number of attributes [36]. Finally, other metrics are introduced as a support to DW testing in [37], namely, *conformity metrics* (to assess how well conformed dimensions have been designed) and *usability metrics* (regarding the navigability of hierarchies).

The metrics proposed in [20] are more relevant to our work since they are related to the schema *correctness* rather than to its understandability/legibility. They are *completeness*, *minimality*, *correct aggregations*, and *minimal sparsity*. Minimality refers to non-redundant elements, correct aggregation to is equivalent to summarizability [38], and minimal sparsity estimates the percentage of empty cells in the cube. It is also checked that schemata are in *multidimensional normal form* [39].

All the approaches mentioned above aim at introducing metrics to measure specific quality attributes of cube schemata. However, they cannot be used to guide decision-makers when inspecting and correcting a cube schema so as to expedite and improve requirements elicitation. Indeed, requirements elicitation asks for a set of guidelines on multidimensional modeling issues, which cannot be expressed by measurable objective metrics —especially within a collaborative setting like the one we consider, where different decision-makers are involved and often disagree on requirements.

2.3. Quality attributes in crowdsourcing systems

The seminal work in [40] provides a categorization of quality attributes. The authors recognise four main groups: intrinsic, contextual, representational, and accessibility. Besides the quality attributes issued from classical information systems, such as precision and accuracy, they identify some new quality attributes exclusively related to the volunteer character of crowdsourcing systems: credibility and attractiveness. Credibility characterises the reputation of the volunteer, while attractiveness defines the popularity of a crowdsourced element. The definition and management of user reputation in the context of crowdsourcing systems has been studied by many (see, for example, [41, 42]). Efforts to link the quality of crowdsourced data to the ability and reputation of its contributors have been carried out in the context of Volunteered Geographic Information (VGI). Some approaches define the trust attributed to the data itself and the reputation of contributors, calculated by investigating the history of edits of geographic features [43, 44]. Specific studies investigate the relationship between the quality of VGI and the number of users editing [45, 46, 47]. In [48] the reputation of contributors is used to develop a rule-based system that identifies potential instances of vandalism in the OpenStreetMap (OSM) database. Following the work in [47], the authors of [49] develop a VGI evaluation model for features trustworthiness and user reputation as a data quality proxy measure. An extension of this model is presented in [50] and applied to the assessment of the quality of street semantics in OSM.

Inspired by some of this work, in this paper we provide our definition of volunteer reputation. In many of the approaches presented in the literature a user reputation depends on many factors, including the previous behaviour of that person [51]. Here, by behaviour we intend the set of interactions a user has with the system. In the context of schema definition, we do not have a historic set of defined cube schemata since, once the DW is implemented, then no further cube schemata are defined or modified. However, we can use the history of data contribution of volunteers in order to assess the reputation of users. For example, this could include any data they added, any edits they made to data inputted by others, etc. Indeed, the historic data contribution of volunteers could indicate whether they are involved in data vandalism (where vandalism is detected by quality control measures), and consequently whether they are likely to carry out vandalism in defining cube schemata for such data. It could also indicate they are trusted by other users in case where other users access their data but do not apply any edits to them.

2.4. Collaborative process recommendation

Group Decision Support Systems (GDSSs) have been defined as "interactive computer-based environments that support concerted and coordinated team effort towards completion of joint tasks" [52]. GDSSs have explicit and implicit advantages that must be understood by the responsible leaders to reach the best of their potential team-oriented productivity [53]. The explicit benefits are mainly, but not exclusively, a better problem definition, more group cohesiveness, a higher number of solutions with better quality, and stronger team commitment to those solutions' adoption and implementation. The implicit ones include a remarkable reduction of the staff's engagement time to reach final decisions and budget savings thanks to the boosted productivity. The original purpose of GDSSs also includes the exploitation of opportunities that information technology tools can offer to support group work.

Many studies have evaluated GDSSs, and showed that they can improve the productivity by increasing the information flow between participants, by generating a more objective evaluation of information, by improving synergy inside the group, by saving time, etc. [54]. The specific use of GDSS systems for requirements elicitation has been discussed since the early 1990s [55, 56, 57]. For example, in [58] a methodology is proposed called Decision Support in Software Engineering (SEDS), which suggests presenting the perspectives of various stakeholders for a better definition of requirements. Also, among the others, in [59] an approach is introduced that helps, in a corporate environment, to manage requirement elicitation. Specifically, some studies highlighted that the efficiency of using GDSSs depends strongly on the presence of a facilitator [60]. Group facilitation is defined as a process in which a person, who is considered as trustworthy by all the group members, intervenes to help improve the way they identify and solve problems, and make decisions [61]. Though in general the role of the facilitator in collaborative process recommendation is very complex and high expertise is required, using a GDSS to support collaboration makes this role much simpler since the decision algorithms are automated.

Recommendation systems can be an important asset in the absence of a professional facilitator. This is the reason why we make use of recommendation to tailor our methodological approach. When a system suggests several items to its users, the recommendation challenge becomes how to provide the most convenient set of propositions tailored to each user's needs. This has been the focus of many commercial and academic research approaches [62]. However, since this is not the main contribution of our approach, we adopt a simple question-based recommendation [63] as a proof of concept that more advanced techniques can be incorporated for further improvements.

3. Working example

Before we describe the working example we will use along the paper, we briefly summarise the basic concepts of the multidimensional model, used to organize data in DWs. A multidimensional schema is based on the concepts of *cube* (analysis subject), *dimensions* (axes for analyzing a cube), and *measures* (numerical attributes of cubes). A cube can have several dimensions and measures. A dimension can be organised in *hierarchies*, each composed of a set of *levels* expressing different possible aggregations of that dimension. *Indicators* are described by a measure and an aggregation function, used to aggregate measure values along hierarchies.

Our working example is related to a real case study concerning the analysis of biodiversity in the agricultural context, and derives from the French ANR project VGI4Bio (`www.vgi4bio.fr/`). Data are obtained from *Faune-Aquitaine* (`faune-aquitaine.org`), an online observation notebook for naturalists which enables them to archive and consult their own observations. This is also a tool to share data and to improve knowledge about the regional fauna, thanks to the pooling of observations from thousands of field naturalists. To ensure data reliability, a validation process is constantly carried out. First of all, incoming data are automatically scanned to check that the observations are compatible with the basic knowledge about species biology (mainly phenology, behaviour, consistency, and distribution). Data are also manually checked by some validators, who are expert naturalists with access to additional functionalities. Validators

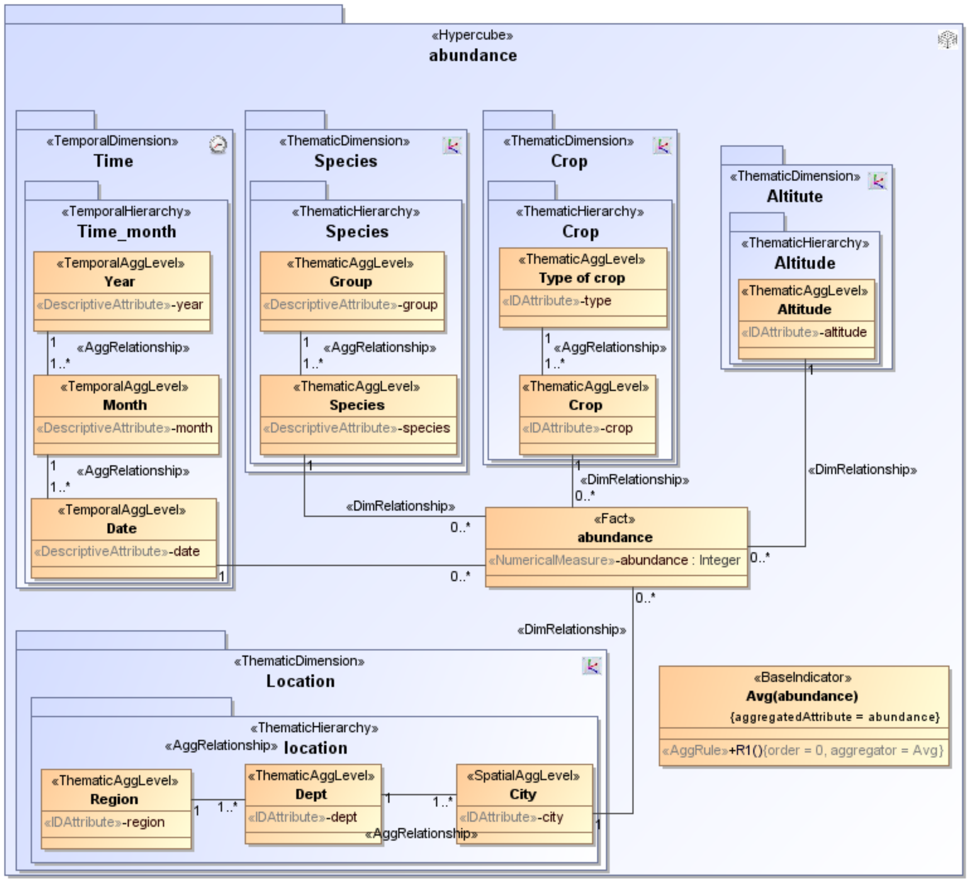

Figure 1: Cube schema for agro-biodiversity analyses; the schema is called abundance as it describes the abundance of species

can tag questionable observations and discuss with their authors to improve data quality. Outliers are also identified a-posteriori, when datasets are used to create distribution maps or phenological analyses.

In this work we focus on the cube schema in Figure 1, which is called abundance (as it describes the abundance of species) and was designed starting from the requirements expressed by volunteers. The schema is represented using the ICSOLAP UML profile [64]. The abundance cube has six dimensions:

- Crop, which includes levels Crop and Type of crop (the latter describing the group of crops used in the plot, e.g., wheat).
- Species, whose hierarchy includes two levels describing the species and its group (for instance, species 'eagle' belongs to the 'Accipitridae' group).
- Location, whose hierarchy includes three levels modeling cities, depart-

ments, and regions, respectively.

- Time, with one hierarchy that groups dates into months and years.
- Altitude, describing the altitude at which each observation was made.

The abundance cube has one indicator, Avg(abundance), which measures the average abundance of each species as observed on each date by each user for each type of crop. Thus, this cube enables analysts to monitor the evolution of biodiversity over time and space, and by type of crop.

4. A profile-aware methodological framework

In this section we sketch a general profile-aware framework for multidimensional modeling methodologies.

We start by recognizing that, in most DW projects, the users involved can be classified into two groups based on their *authoritativeness*, which depends on their self-confidence on the application domain and on the analysis requirements, as well as on how likely it is they will be respected and obeyed in the context of the DW project:

- We call *end-users* the non-authoritative users, i.e., those who are supposed to have a good knowledge of the application domain, will use the DW for their analysis, may contribute to the project by expressing some requirements that are not guaranteed to actually be taken into account faithfully.
- We call *decision-makers* the authoritative users, i.e., those who have an excellent knowledge of the application domain, will use the DW for their analysis, can validate the end-users requirements and integrate them their own requirements.

Note that in a typical DW project taking place within a company, decision-makers correspond to the so-called *key users*, i.e., those who have a deeper knowledge of the application domain and/or who play a leading role in the company. End-users are those who do not have a broad picture of the company strategy and will mainly use the DW for routine analysis. Conversely, in citizen science projects, end-users correspond to volunteers, who express their preliminary and "local" analysis needs, while decision-makers correspond to so-called *committers*, i.e., authoritative experts of the application domain who collaboratively validate and integrate the volunteers proposals [7].

We further specialise both end-users and decision-makers:

- For end-users we consider their level of *engagement* in the project: we distinguish between projects whose end-users are engaged (mostly, company projects) and projects whose end-users are volunteers (citizen science projects).

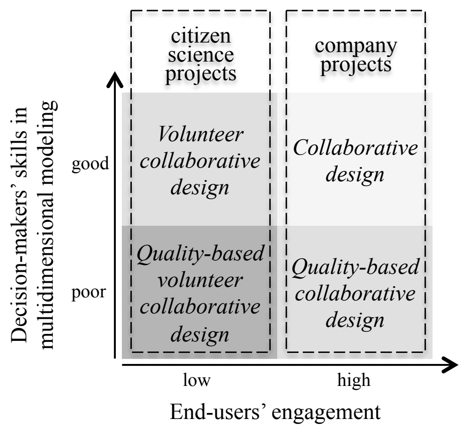

Figure 2: Magic quadrant for DW projects and multidimensional modeling methodologies

- For decision-makers we consider their skills in multidimensional modeling, which can make them more or less autonomous and confident in validating and integrating the analysis requirements.

As a result, we can draw some sort of "magic quadrant" for DW projects as shown in Figure 2 where, with respect to the horizontal axis (end-users' coordinate):

1. The two right-most quadrants cover typical company projects, where end-users are strongly engaged. Thus, there is no need to assess their reputation; in other words, we assume that the risk that they may maliciously influence the project to achieve some personal goal (e.g., by removing dimensions/indicators that can be explanatory for sensible phenomena such as treatments for agricultural biodiversity) is negligible.

2. The two left-most quadrants cover citizen science projects, where end-users are volunteers (i.e., they are not paid for the task they execute in the context of the project and are likely to be loosely engaged). In this case, assessing the reputation of end-users is crucial to ensure that the work made by volunteers is reliable, as they might make mistakes due to their lack of expertise or they might deliberately want to perform vandalism [48].

With respect to the decision-makers' coordinate (vertical axis):

1. If decision-makers have sufficient skills in multidimensional modeling (two top quadrants), a basic collaborative approach can be pursued to coordinate all users and obtain comprehensive cube schemata.

2. Otherwise, decision-makers can hardly evaluate the quality of the cube schemata proposed by end-users [25]. In particular, they are likely unable to identify typical problems of cube schemata such as redundancy,

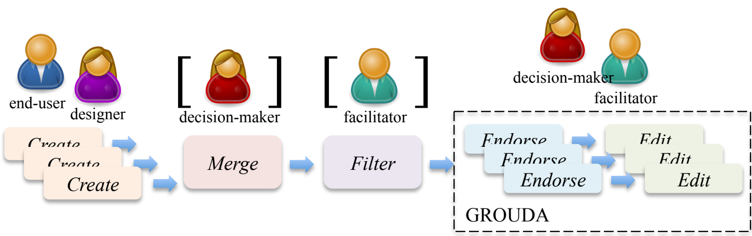

Figure 3: Methodological framework

Time	Crop	Location		Avg(abundance)
Time	*Crop*	*Location*		
Year	Type of Crop	Dept	Location	
2019	Wheat	Allier	Aubiere	14
2019	Wheat	Allier	Clermont-Fd	12
2020	Apple	Cantal	Issoire	15

Figure 4: A sample pivot table in Excel

completeness, etc. Besides, integrating these schemata also taking into account their own requirements will be very difficult for them. Thus, some support and guideline must be given to them for these tasks to achieve a quality-based collaborative design (two bottom quadrants).

Based on the above discussion, we aim at devising a profile-aware methodological framework that can flexibly adapt itself to cover all four quadrants in Figure 2. To this end we extend and generalise the methodology proposed in [7], which falls in the top-left quadrant (volunteer collaborative design). In the following we sketch the general methodological framework, while in Section 7 we specifically describe how it is tailored to the most challenging scenario, i.e., that of non-engaged end-users and decision-makers not skilled in multidimensional modeling (bottom-left quadrant of Figure 2).

Besides decision-makers and end-users, our framework relies on two more actors, corresponding to DW experts in [7], both skilled in multidimensional design but with little or no knowledge of the application domain:

- *designers*, who support end-users in expressing their requirements, and
- *facilitators*, who organise the collaboration of decision-makers and select the cube schema elements to be validated by them.

The main steps of the methodological framework we propose are shown in Figure 3 and briefly commented below:

1. During the first step, *Create*, end-users define their own cubes supported by designers. Specifically, as done in [7], the requirements of each end-user are captured in five steps: (i) a goal model is created by a designer by interviewing the end-user; (ii) the end-user expresses detailed requirements for each goal by drawing pivot tables in Excel; (iii) these requirements are

refined by the designer via semi-structured interviews with the end-user; (iv) each pivot table is automatically translated into an ICSOLAP cube schema; and (v) a prototype is automatically obtained from that cube schema using the ProtOLAP tool [65], to let the end-user validate her analysis requirements. As to (ii) we observe that, while other approaches have been devised in the literature to capture requirements (e.g., based on MDX expression or SQL queries), using pivot tables is preferred in case of users who are unskilled in multidimensional modeling, so they cannot express their analysis needs directly in terms of multidimensional concepts such as measures and dimensions, and even in terms of classical requirements engineering concepts such as goals, KPIs, etc. [25]. Conversely, these users can easily understand and validate the results of prototype implementations, which normally consist of pivot tables returned by OLAP tools. As to (iv), the complete algorithm for translating pivot tables into ICSOLAP schemata is shown in [25]; the basic idea is to map the headers of each pivot table into multidimensional concepts (namely hierarchies, dimensions, levels, and indicators).

Example 1. *Figure 4 shows a pivot table drawn by an end-user and including three dimensions and one indicator; Figures 5 and 6 show the corresponding ICSOLAP schema and the prototype created using the JRubik OLAP client.*

2. The previous step produces a set of ICSOLAP schemata, obtained from the requirements the different end-users. During the second step, *Merge*, these schemata are semi-automatically merged to generate one or more integrated cube schemata. Indeed, using pivot tables encourage decision-makers to think in terms of single queries rather than in terms of a global multidimensional schema, so an integration step is mandatory. In terms of DW design, this corresponds to the fusion of a set of multidimensional schemata, for which some approaches have been proposed in the literature. We adopt the one proposed in [7], in turn inspired by [66, 67]. Basically, for each common hierarchy (i.e. hierarchy with the same name), we fuse the levels present in the different schemata by preserving their functional dependencies, so that the hierarchies designed by a user may be enriched with levels designed by other users. Standard semantic reconciliation techniques (e.g., [68]) can be adopted to automate the derivation of inter-level matchings. The contradictions possibly arising in hierarchies are solved manually by the decision-makers.

Example 2. *Consider again the cube schema in Figure 5, to be merged with the one in Figure 7. Two different versions of the* Location *hierarchy are found, one including levels* Region *and* Location*, one including levels* Dept *and* Location*. Therefore, the designer has to decide whether two separate hierarchies should be created or these three levels should be included in the same hierarchy. In this case, a many-to-one association*

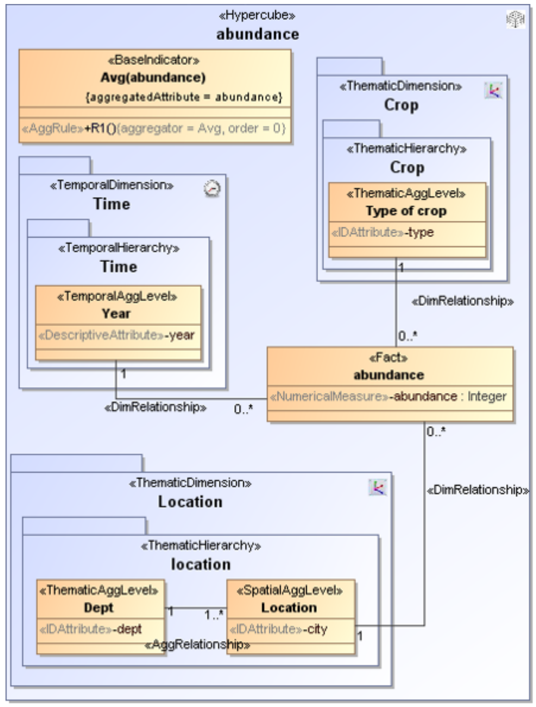

Figure 5: ICSOLAP schema obtained from the pivot table of Figure 4

holds between Dept *and* Region*, so she chooses the second option; the merged ICSOLAP schema is shown in Figure 8.*

3. The integrated cube schemata may include a large number of elements; the *Filter* step aims at reducing the validation effort of decision-makers by letting them focus only on the controversial elements, which are more likely to give rise to discussions and conflicts. This can be done by a facilitator if she has some knowledge of the application domain. Otherwise, filtering must be automated; to this end we introduce a metric aimed at assessing how trustworthy each schema element is, also considering the reliability of volunteer end-users (Section 5).

4. Each controversial element is then collaboratively validated by decision-makers (*Endorse*). As a consequence, some changes to the cube schema (e.g., delete a level or rename a hierarchy) may be required (*Edit*). Both activities are supported by the GROUDA GDSS (see Section 8) and coordinated by a facilitator. To cope with the case in which decision-makers have no multidimensional modeling skills, we introduce a set of quality attributes aimed at supporting them in inspecting cube schema and detecting errors (Section 6).

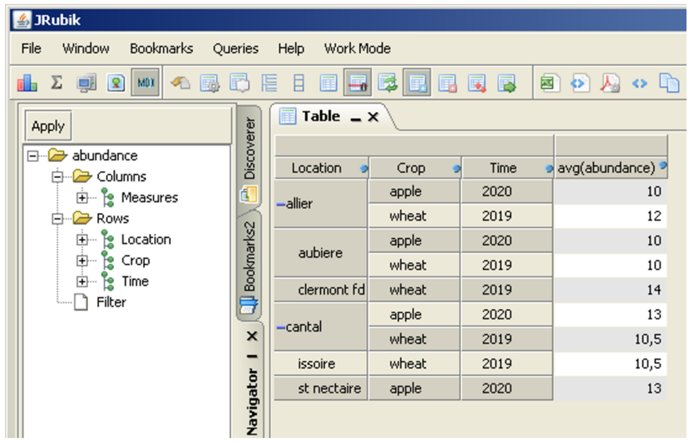

Figure 6: JRubik prototype for the ICSOLAP schema of Figure 5

To complete the design of the DW, the resulting ICSOLAP cube schema is automatically prototyped and validated as described in [7].

5. A metric for filtering

In this section we introduce a *trustworthiness* metric, inspired by the literature on crowdsourcing systems, aimed at assessing the reputation and reliability of end-users (specifically, volunteers in the two left-most quadrants of Figure 2) and, therefore, of the cube schema they contribute. Trustworthiness represents the degree of confidence that can be associated to each level and each indicator suggested by end-users; it is based on three factors, namely, *expertise*, *volunteer reputation*, and *attractiveness*. Importantly, trustworthiness will be used during the *Filter* step to relieve decision-makers of the endorsement effort for non-controversial elements of cube schemata.

1. *Expertise*. Each end-user indicates her degree of expertise concerning the application domain of the cube schema using the ordinal scale {'low'; 'medium'; 'high'; 'expert'}. The expertise of end-user i, indicated Exp_i, is then obtained by mapping these ordinal values to numerical values (0.3, 0.5, 0.7, and 1, respectively). For instance, in our working example, ecology and agronomy researchers would classify themselves as 'expert', a farmer as 'medium' (since she only has empirical knowledge about agrobiodiversity), while a naturalist amateur as 'low' (since she has no knowledge about agriculture).

2. *Volunteer reputation*. Reputation belongs to the community, not to the person whose trust was evaluated and it depends on many factors, including the previous behaviour of that person [51]. In a crowdsourcing project, where members of the community are able to accept/reject/edit

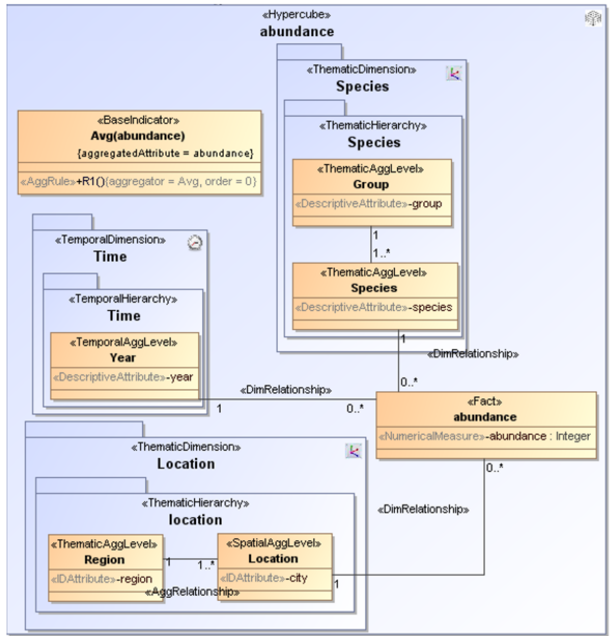

Figure 7: An alternative ICSOLAP schema to be merged with the one in Figure 5

other people's contributions, the reputation of volunteer i can be seen as a function of the ratio of correct data (or data accepted by the community) and all the data that i contributed. More formally, we define the reputation of volunteer i as:

$$Rep_i = f(1 - \frac{rd_i}{d_i})$$

where:

- rd_i represents the number of rejected data entries made by i;
- d_i represents the number of all data entries entered by i (therefore $rd_i \leq d_i$);
- f is a function with domain and codomain $[0, \ldots, 1]$.

Calculating the reputation of end-users based on the quality of the data they provide gives us a means of adjusting the self-assessment provided by users when indicating their expertise. Indeed, expertise is taken into account when calculating the trustworthiness of multidimensional elements, which we consider as a proxy for their quality, to identify elements that require further checks. Therefore, it is important to have a mechanism that can capture the potential for vandalism and/or for incorrect entries made by non-expert users when calculating trustworthiness.

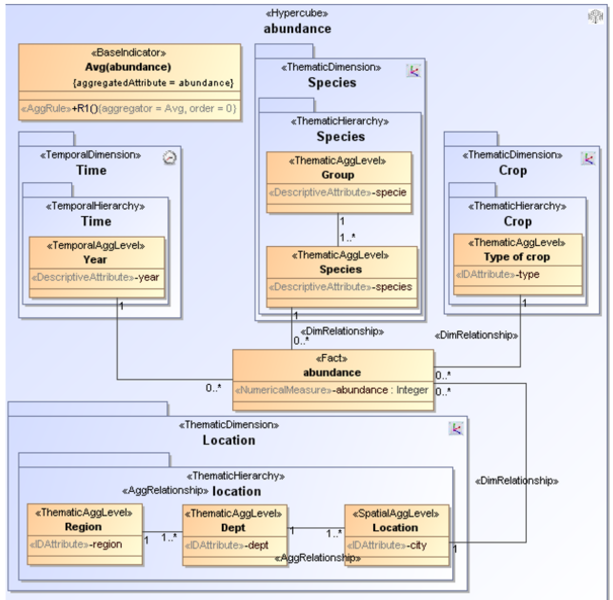

Figure 8: ICSOLAP schema obtained by merging those in Figures 5 and 7

3. *Attractiveness* represents the "popularity" of a multidimensional element within the proposed cube schemata, and provides an important measure of its quality. Indeed, according to the *many eyes principle* [69], the more a piece of data is defined/edited, the higher its quality. Formally, given an element e and end-user i, we define

$$Attr_i(e) = \begin{cases} 1 & \text{, if } i \text{ has defined } e \\ 0 & \text{, otherwise} \end{cases}$$

For instance, if five end-users agree on the presence of the Crop dimension, its attractiveness is 1 for all these end-users.

4. *Trustworthiness.* We can now define the trustworthiness of a multidimensional element e as:

$$Trust(e) = \frac{\sum_i (w^{exp} \times Exp_i + w^{rep} \times Rep_i) \times Attr_i(e)}{\sum_i (w^{exp} \times Exp_i + w^{rep} \times Rep_i)} \tag{1}$$

where w^{exp} and w^{rep} are weights such that $w^{exp} + w^{rep} = 1$. Therefore, trustworthiness is a value between 0 and 1, where 1 means that all end-users defined this element. As previously mentioned, combining the (self-defined) expertise of end-users and their (quality-dependant) reputation in the weighted sum allows to moderate the effect the expertise has in assessing the trustworthiness of elements. This aims at avoiding that either malicious volunteers identify themselves as experts for vandalism actions, or non-expert users do the same with no malicious intention but still resulting in several incorrect contributions. Similarly, if an expert end-user is too modest in her self-assessment, her high reputation will compensate for it. The values of w^{exp} and w^{rep} can vary depending on the particular scenario. For example, in a quite controlled environment, where it is very unlikely to have malicious users and the self-assessment is reliable (e.g., because very specific guidelines are provided), both weights could have the same value. In other cases, more weight could be given to the reputation vs. self-assessed expertise. Where the self-assessment is totally unreliable, it could be disregarded altogether.

To give a theoretical validation of the trustworthiness metric, we adopt the DISTANCE framework [70]. Given a set of end-users $\mathcal{U}$ and a multidimensional element e, our reference universe $\mathcal{S}$ is the set of all possible project situations where any subset of end-users has defined element e. The steps of the validation process are then solved as follows:

1. **Find measurement abstraction**: we adopt function $abs : \mathcal{S} \rightarrow 2^{\mathcal{U}}$ such that $abs(s) = U_s$, where $U_s \subseteq \mathcal{U}$ is the set of end-users who defined element e in project situation $s \in \mathcal{S}$.

2. **Model distances between measurement abstractions**: Since measurement abstractions are sets of end-users, only two elementary transformations are required: one for adding an end-user to a set and one for removing an end-user from a set.

3. **Quantify distances between measurement abstractions**: let $Trust_U(e)$ denote the value taken by Formula 1 when evaluated with reference to a project situation where U is the set of end-users who defined e (note that U impacts on the values of $Attr_i(e)$ for each i). Given two project situations s and s', we measure their distance as

$$\delta(s, s') = Trust_{abs(s)-abs(s')}(e) + Trust_{abs(s')-abs(s)}(e)$$

   It is straightforward to prove that δ satisfies identity of indiscernibles, symmetry, and triangle inequality, so it is a metric[2].

4. **Find a reference abstraction**: the project situation s^* where *no end-user* defined element e, i.e., where $abs(s^*) = \varnothing$.

5. **Define the software measure**:

$$\begin{aligned}\delta(s, s^*) &= Trust_{abs(s)-abs(s^*)}(e) + Trust_{abs(s^*)-abs(s)}(e)\\ &= Trust_{U_s}(e)\end{aligned}$$

[2]It has been proven that the symmetric difference model can always be used to define a metric when the set of measurement abstractions is a power set like in our case [71]. Though the function we use here does not simply count the number of end-users in each set, it weighs the presence of each end-user with a constant that depends only on that end-user, so all properties of metrics are preserved.

Table 1: Quality attributes for cube schemata and their applicability to cube schema elements

	Indicator	Set of indicators	Level	Set of levels	Hierarchy	Set of hierarchies	Dimension	Set of dimensions	Cube
Completeness		yes		yes		yes		yes	
Precision	yes						yes		
Relevance	yes		yes						
Minimality		yes		yes		yes		yes	
Consistency	yes				yes				
Certainty	yes		yes		yes		yes		
Confidentiality	yes		yes						
Usability									yes

6. Quality attributes for endorsement

Inspired by the quality attributes defined for databases, the attributes we introduce in this section aim at supporting decision-makers not skilled in multidimensional modeling and are intended to be adopted in scenarios corresponding to the two bottom quadrants of Figure 2.

Specifically, these quality attributes are used to support decision-makers in detecting possible errors in multidimensional modeling during the *Endorse* step. Note that they are not formally defined as metrics because (i) they can be evaluated only ex-post, i.e., after the design process is completed and the final version of a cube schema is available; (ii) they are subjective and multi-valued, i.e., each decision-maker may give different evaluations for these attributes and may be difficult to reach an agreement; and (iii) they are applied to fine-grained parts of a cube schema and are defined as Booleans. Indeed, they are not meant to *measure* the quality of schemata, but to provide a sort of *checklist* that each decision-maker is encouraged to inspect on all different parts of a cube schema to ensure that all quality aspects are properly considered.

The applicability of quality attributes to cube schema elements is summarised in Table 1 and detailed below. Some examples are referred to the cube schema in Figure 9, obtained by merging the schemata proposed by different end-users, where some errors are present.

1. *Completeness* refers to sets of multidimensional elements, and indicates whether all necessary concepts have been modeled in the cube schema. It can be applied to the following sets of elements:

    - *set of indicators* (e.g., the median and the standard deviation of abundance might be useful as well, besides its average);
    - *set of levels of a hierarchy* (e.g., the Location hierarchy might be considered incomplete should an additional Plot level be useful for farmers' analyses);

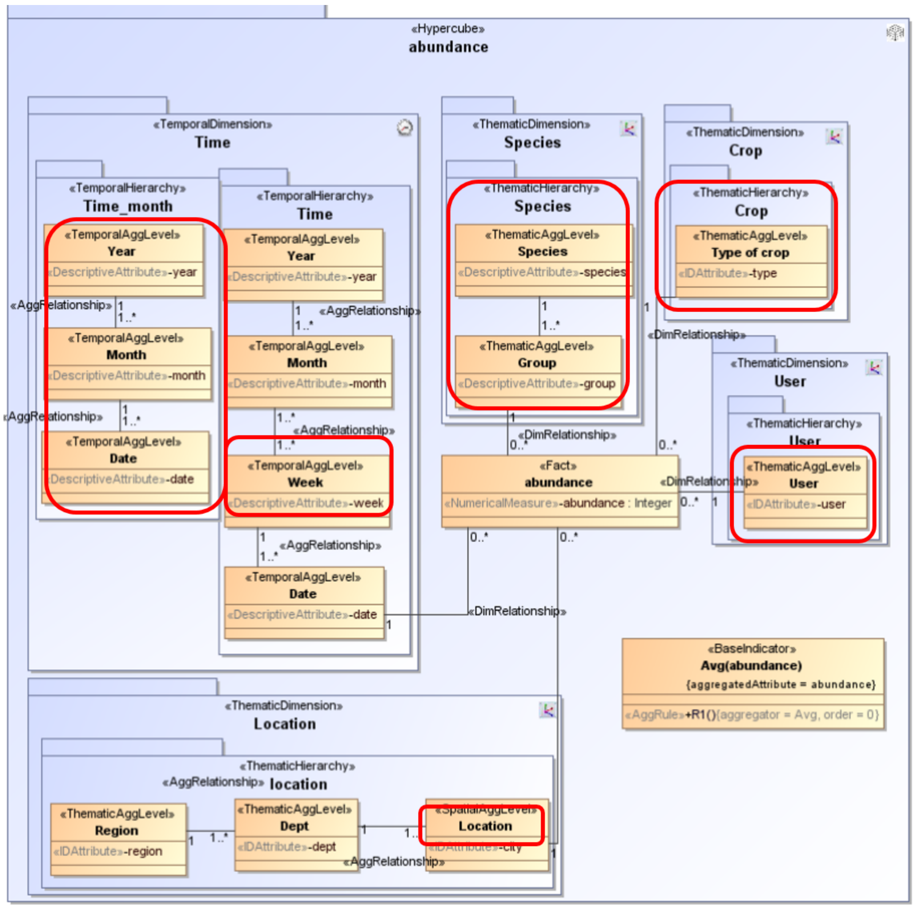

Figure 9: Cube schema for agro-biodiversity analyses with some errors (in red)

- *set of hierarchies* (e.g., a hierarchy for classifying species according to some specific taxonomy might be missing);
- *set of dimensions* (e.g., a dimension representing altitudes is mandatory to characterise the abnormal presence of some species).

2. *Precision* refers to single multidimensional elements, and indicates whether an element is represented with sufficient precision and detail. It can be applied to two types of elements:

   - *indicator* (for instance, using an integer to measure the average abundance is not precise enough since values are very small —typically from 1 to 15; therefore, a higher precision with two decimal digits is required);
   - *dimension* (e.g., biodiversity is strongly related to specific crops, thus,

the Crop dimension —whose finest level in Figure 9 is Type of crop— is not precise enough and level Crop should be added).

3. *Relevance* refers to single multidimensional elements and indicates whether they are useful for analysis; it can be applied to two types of elements:
   - *indicator* (e.g., an indicator based on an irrelevant measure might have been created);
   - *level* (e.g., the User and Week levels in Figure 9 are not relevant for the analysis of biodiversity).
4. *Minimality* refers to sets of multidimensional elements and indicates whether they present redundancies (i.e., duplicates). It can be applied to different sets of schema elements: *set of dimensions*; *set of indicators*; *set of levels of a hierarchy*; and *set of hierarchies of a dimension* (e.g., hierarchy Time_month in Figure 9 is redundant because hierarchy Time already contains all the levels of Time_month).
5. *Consistency* refers to single multidimensional elements and indicates to what extent the rules characterizing the application domain have been adhered to. It can be applied to two types of elements:
   - *indicator* (e.g., an indicator might use an inappropriate aggregation operator);
   - *hierarchy* (e.g., in the Species hierarchy of Figure 9, groups of species are children of species, while it should be the opposite).
6. *Certainty* refers to single multidimensional elements and indicates that there are no ambiguities in the names chosen for them. It can be applied to four types of elements: *hierarchy*, *dimension*, *indicator*, and *level* (e.g., the Location level in Figure 9 is ambiguous since it apparently represents the exact geographical coordinates of an observation, while it actually represents the city where the observation was made).
7. *Confidentiality* is inspired from [40] and indicates whether a multidimensional element can be part of the cube schema or it should be removed for legal, anonymization, or confidentiality issues. It can be applied to two types of schema elements, namely, *indicator* and *level* (e.g., the User level in Figure 9 may cause privacy and confidentiality problems).
8. *Usability* indicates if, overall, a cube schema facilitates analysis. For example, a cube with 30 dimensions can lead to unreadable pivot tables in OLAP clients.

7. Quality-based volunteer collaborative design

In this section, we explain in detail how the profile-aware framework introduced in Section 4 is instanced into a specific methodology for the scenario where end-users are volunteers and decision-makers have poor skills in multi-dimensional design (bottom-left quadrant of Figure 2). To address the most challenging case, we also assume that the facilitator has a poor knowledge of the application domain, so that the Filter step must be automated.

Steps *Create* and *Merge* are the same as in [7]; the other steps are described in the following subsections.

7.1. Filter

The goal of this step is to reduce the validation effort of decision-makers during the next step. This is done by forcedly validating *a priori* all non-controversial multidimensional elements. This can be done by a facilitator if she has some knowledge of the application domain as well as sufficient time, otherwise it can be done automatically as described below.

The main idea is that, when end-users are reliable and they agree on the definition of a multidimensional element, then that element is well-defined. This is expressed by the trustworthiness metric defined in Section 5. The multidimensional elements whose trustworthiness is above a given threshold are automatically validated, thus saving the decision-makers' effort in voting those elements during the *Endorse* step.

For instance, a facilitator with good skills in ecology who also participated in the collection of data can set a relatively high threshold, e.g., 80%. Indeed, during the *Endorse* step it will be easy for her to manually validate controversial elements such as location, since she is aware that the name Location used in the application for data collection actually refers to city.

7.2. Endorse

During this step, the multidimensional elements that were not validated during *Filter* are checked by decision-makers to see if they can be validated or not. Should decision-makers have good skills in multidimensional modeling, they could quickly do this by just discussing each element during the required meetings conducted by the facilitator. In the scenario we are considering, the decision-makers' skills in multidimensional modeling are poor, so we propose to rely on the quality attributes described in Section 6.

The endorsement process takes place as follows. First of all, the facilitator explains the quality attributes to the decision makers supported by some examples. Then, for each multidimensional element, each applicable quality attribute is evaluated. The multidimensional elements are examined following a bottom-up approach, i.e., from the finest ones to the coarsest ones. More precisely, considering the part-of relationships between elements, indicators and levels are considered first, followed by hierarchies, then by dimensions, and finally by the whole cube. The whole process is sketched in Figure 10.

Example 3. *For instance, after checking levels* Date*,* Month*, and* Year *for relevance, certainty, and confidentiality, the* Time_month *hierarchy is checked for certainty and consistency. After the set of hierarchies of the* Time *dimensions*

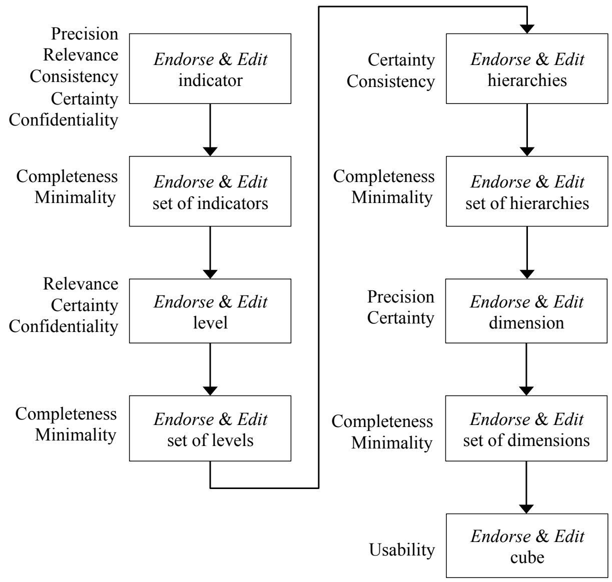

Figure 10: Breakdown of the *Endorse* step

has been checked for completeness and minimality, and finally the Time *dimension is checked for precision and certainty.*

The check of each quality attribute on each multidimensional element is done collaboratively through a majority vote. We choose a simple majority vote, without taking into account the expertise of each decision-maker, because multidisciplinary skills are involved in the endorsement of a cube schema. In fact, to enable an expertise-weighted vote, the decision-makers would have to define their expertise for each multidimensional element, which would make the *Endorse* step too time-consuming. For example, checking the Crop dimension requires agronomy skills, while Species should be evaluated by a naturalist. To solve this issue, we give each decision-maker the possibility of "not voting" when she considers herself not skilled enough to express her opinion on a specific multidimensional element. This may lead to a neutral result of the voting process, which must then be solved by decision-makers via a free discussion.

Example 4. *In our working example, the* User *level was voted as relevant only by two decision-makers, while the other three thought it was irrelevant. Thus,* User *was deleted during the* Edit *step (as described in Section 7.3). Conversely, the* Location *level was voted as relevant by all decision-makers, so it was checked for certainty. Four decision-makers out of five voted* Location *as not certain (i.e., they deemed its name as ambiguous), so this level was renamed during the* Edit *step.*

We close this subsection by observing that, as depicted in Figure 10, the endorsement of each element is immediately followed by its editing. Since ele-

Table 2: Actions on cube schema elements that can be triggered in presence of quality problems during *Endorse*

	Indicator	Level	Hierarchy	Dimension
Completeness	add	add	add	add
Precision	modify	add		
Relevance	delete	delete		
Minimality	delete	delete	delete	delete
Consistency	modify		modify	
Certainty	rename	rename	rename	rename
Confidentiality	delete	delete		
Usability	delete	delete	delete	delete

ments are examined from the finest ones to the coarsest ones, performing the *Edit* step on a component element may affect the subsequent endorsement of the composed element. For instance, if the only hierarchy in a dimension is voted to be redundant (i.e., non-minimal) and is deleted, that dimension will then be empty so it will not have to be endorsed.

7.3. Edit

Once all quality attributes have been evaluated on an element or set of elements, some actions may have to be applied so as to correct the cube schema. These actions are:

- *Delete*. For example, the User level is removed since it causes confidentiality problems.
- *Add*. For example, the Crop level is added to the Crop hierarchy.
- *Rename*. For example, Location level is renamed as City.
- *Modify*. For example, levels Species and Group are inverted in the Species hierarchy.

The collaborative process for this step is implemented as a free discussion among the decision-makers. Table 2 shows which actions on cube schema elements can be triggered in relationship to each quality problem possibly emerged during the *Endorse* step.

8. Implementation

The technological stack for our implementation relies on three components: (i) GROUDA, a GDSS; (ii) PostgreSQL (`www.postgresql.org/`), a well-known relational database management system; and (iii) Mondrian (`mondrian.pentaho.com/`), an open-source OLAP Server.

GROUDA (GROUp Decision & DAta-warehouse) allows decision-makers to organise a variety of thinklet-based group activities, coordinated by a facilitator. A *thinklet* is a "scripted collaborative activity that gives rise to a known pattern of collaboration among people working together towards a goal" [72]. Every

```xml
  <?xml version="1.0"?>
- <Schema name="abundance">
  - <Cube name="abundance">
      <Table name="fact"/>
    - <Dimension name="location" foreignKey="location_id">
      - <Hierarchy primaryKey="location_id" allMemberName="All location" hasAll="true">
          <Table name="location"/>
          <Level name="Region" Trustworthiness="82" type="String" uniqueMembers="true" column="location_region"/>
          <Level name="Dept" Trustworthiness="61" type="String" uniqueMembers="true" column="location_department"/>
          <Level name="Location" Trustworthiness="22" type="String" uniqueMembers="true" column="location_municipality"/>
        </Hierarchy>
      </Dimension>
    + <Dimension name="species" foreignKey="species_id">
    + <Dimension name="Time" foreignKey="date_id">
    + <Dimension name="User" foreignKey="User_id">
    + <Dimension name="Crop" foreignKey="Crop_id">
      <Measure name="AVG(abundance)" Trustworthiness="100" column="abundance" visible="true" aggregator="avg"/>
    </Cube>
  </Schema>
```

Figure 11: An excerpt of the XML Mondrian file for the abundance cube schema

thinklet is composed of three group activity stimuli that are described independently of the technological support of its implemented solution: (i) the tool, i.e., the hardware and software technology used so the thinklet definition allows reproducibility, (ii) the configuration to give a precise parametrization among the various possible combinations of the used tool settings, and (iii) the script that leads to the sequential progress of the activity execution. GROUDA has been implemented as a web-based application in Python with the Django 2.2.2 framework, which allows the availability of these thinklet-based group activities on the web. All data about users and meetings are stored in Postgres. Importantly, GROUDA provides a completely decentralised GDSS, which enables groups of decision-makers to remotely join their meetings. It is also possible to use some of the offered techniques asynchronously, ensuring flexibility in terms of users' availability.

Mondrian is used to implement the DW prototype starting from an XML file that describes cube schemata and is automatically generated from ICSOLAP schemata as described in [7]. GROUDA comes with a java component that takes in input this file so as to display its multidimensional elements on the user interface for voting or discussing. Note that we have extended the original XML schema of Mondrian with an attribute to represent the trustworthiness for the Level and Measure tags. This attribute is used in the *Filter* step to automate the selection of the multidimensional elements to be validated. Even in case the selection is made manually by the facilitator, trustworthiness values are shown to her in the user interface to make her task easier. An excerpt of the XML file for the abundance cube schema is shown in Figure 11.

In order to take into account the varying users' profiles and skills, our framework recommends a different implementation of the *Filter* and *Endorse* steps depending on the answers given to three questions:

1. "*Do you consider yourself an expert of the application domain?*". This question is directed at the facilitator.

    - If the answer is no, the *Filter* step is automated as described in Section 7.1, based on the trustworthiness metric.

- If the answer is yes, the facilitator can manually select the well-defined elements of the cube schemata proposed by end-users.

2. "*Do you think your skills in multidimensional modeling are good?*". This question is directed at decision-makers.

   - If the answer is no for at least one of them, decision-makers must be supported by quality attributes and the *Endorse* step is based on majority votes, as described in Section 7.2.
   - If the answer is yes for all, considering that the evaluation of quality attributes can be long and tedious, the *Endorse* step can be carried out more informally via a free discussion.

3. "*Are the end-users volunteers?*". This question is directed at the facilitator.

   - If the answer is no, the reputation of all end-users is set to 1 when computing trustworthiness.
   - If the answer is yes, the reputation of end-users is computed as discussed in Section 5.

To support the *Endorse* and *Edit* steps we created a thinklet that supports both variants of the collaborative process (voting and free discussion):

- When decision-makers have good skills in multidimensional modeling, the facilitator explains each multidimensional element (Figure 12), invites the decision-makers to discuss them, and annotates their suggestions so as to take the necessary edit actions.

- Otherwise, an additional functionality is shown to enable voting the different quality attributes of multidimensional elements when clicking on each of them (Figure 13). Then, at the facilitator's command, the results of voting are shown to all the participants to keep them informed about the decisions taken (Figure 14).

At the end of the meeting, a report of the results is downloaded by the facilitator who will be in charge of applying the edit actions as discussed by the decision-makers. Remarkably, this process is streamlined and accelerated thanks to the combined use of the ICSOLAP profile and of the ProtOLAP tool. Indeed, each edit action on a cube schema is quickly carried out by the facilitator on the corresponding ICSOLAP schema. Then, ProtOLAP automatically generates a prototype used to continue the process.

9. Experiments

In this section we describe two experiments carried out to put our approach to the test on the agro-biodiversity case study. The first experiment, called

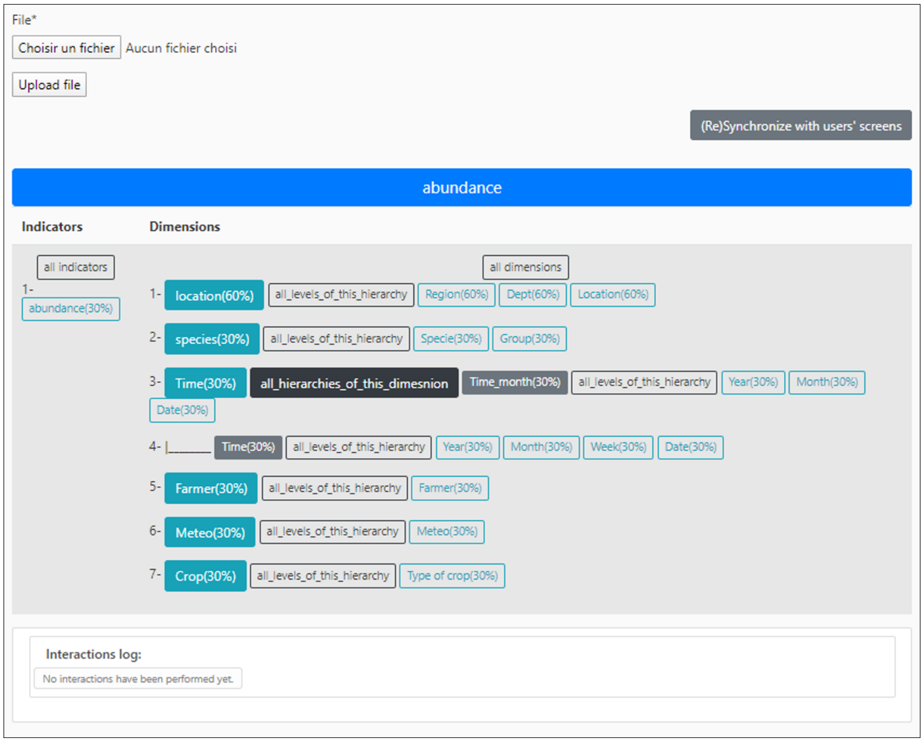

Figure 12: The facilitator summary of multidimensional elements in GROUDA

Exp1 from now on, concerns the `abundance` cube schema, used as a working example throughout the paper; this cube enables analyses of the biodiversity of birds and is fed with the LPO (Ligue pour la Protection des Oiseaux) dataset. The second experiment, **Exp2**, has been conducted on a cube fed with data from the Observatoire Agricole de la Biodiversite (`observatoire-agricole-biodiversite.fr`) and concerning the biodiversity of bees according to the applied treatments at the plot scale.

9.1. Description

In **Exp1**, five volunteers were involved in the definition of cube schemata; they created 14 cube schemata overall, which were then merged into one cube schema (the one in Figure 9).

In **Exp2**, five volunteers defined 12 cube schemata. From them, a single cube schema was obtained with one indicator (the average abundance), a temporal dimension, a treatment dimension, a spatial dimension (grouping plots into farms, cities, departments, and regions), a crop system dimension, and a neighboring land use dimension (e.g., wood, urban area, or other cultivated area). Both experiments were focused on the novel steps proposed in this work, namely, the *Filter* and *Endorse* steps.

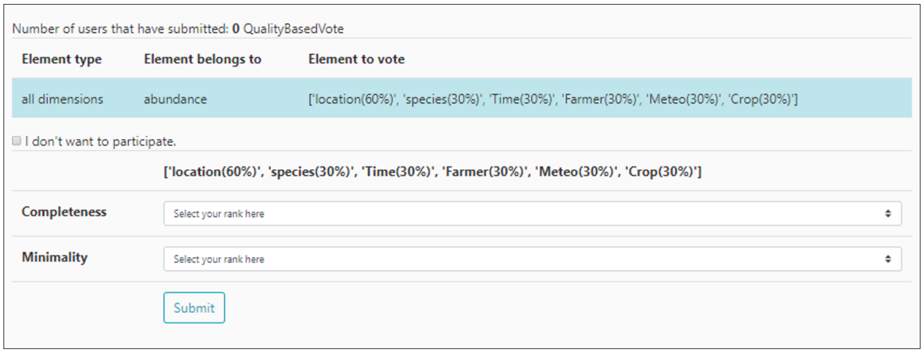

Figure 13: The voting phase in GROUDA

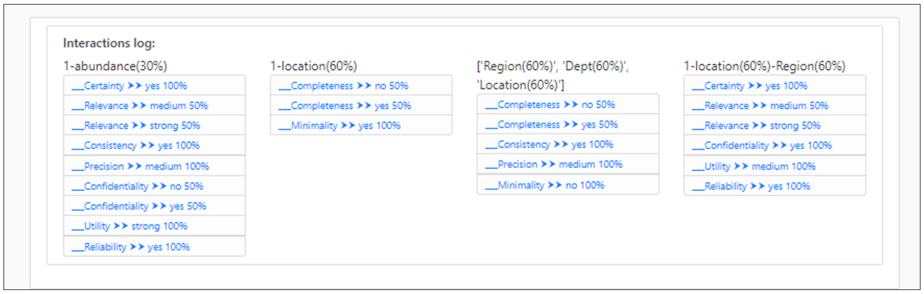

Figure 14: Voting results in GROUDA

To assess the trustworthiness of each multidimensional element during the *Filter* step, we had to measure the reputation of volunteers and the expertise of end-users. Tables 3 and 4 show the reputation of each volunteer involved in **Exp1** and **Exp2**, respectively, together with its components: the number rd of data rejected and number d of data collected. These numbers count the observations of taxa made by each volunteer. These observations are collected in the *Faune-Aquitaine* notebook to be manually accepted (or rejected) by some expert naturalists; note that not all observations are actually loaded in the cubes, since some of them refer to species not monitored. We also observe that the data collected in **Exp2** are less than in **Exp1**, since they are collected by farmers and/or researchers with a more complex and time-consuming protocol [73].

To compute the reputation we used a simple quadratic function:

$$Rep_i = (1 - rd_i/d_i)^2$$

This is because, by statistically analyzing the data, we observed that many of the data contributors provide large amounts of data, and the majority of contributors make very few mistakes. In other words, the ratio rd_i/d_i is in almost all cases very low. However, in some such cases, if a contributor provides a very large amount of data items (e.g., 10000) and makes several mistakes (e.g.,

Table 3: Volunteers' reputation and expertise in **Exp1**

i-th volunteer	rd_i (rejected data)	d_i (collected data)	Rep_i	Exp_i
V1	2	11	0.67	Expert (1)
V2	0	32909	1	Expert (1)
V3	0	14529	1	Low (0.3)
V4	0	2929	1	High (0.7)
V5	23	1621	0.97	High (0.7)

Table 4: Exp2: Volunteers' reputation and expertise in **Exp1**

i-th volunteer	rd_i (rejected data)	d_i (collected data)	Rep_i	Exp_i
V6	3	20	0.72	Expert (1)
V7	0	250	1	Expert (1)
V8	0	150	1	Expert (1)
V9	2	270	0.98	Medium (0.5)
V10	4	350	0.97	Medium (0.5)

100) but still a very small number with respect to her overall contribution, this ratio would be very low, while the number of mistakes made is still quite high. By using a quadratic (or higher order) function we better take this into account. Of course, more complex functions could also be used.

As to the expertise, we observe that the analysis of data requires skills that may be different from those needed to collect the same data. Indeed, in our case study, the collection of data needs basic naturalist skills, while the analysis of these data for agro-biodiversity issues requires ecological knowledge. For instance, volunteer V4 is a good data collector, but he is not an ecological expert, so he evaluated his expertise as 0.7.

To evaluate the effectiveness of the quality attributes for the *Endorse* step in **Exp1**, we focused on the following aspects (see Figure 9):

1. completeness of the set of dimensions;
2. precision of the Crop dimension;
3. relevance of the User level;
4. relevance of the Week level;
5. minimality of the set of hierarchies of the Time dimension;
6. consistency of the Species hierarchy;
7. certainty of the Location level;
8. confidentiality of the User level.

Then, we created two groups of three decision-makers each and let them investigate these problems using:

- for group A, a preliminary analysis made by each decision-maker on its own, and then a global free discussion;

- for group B, the majority vote based on quality attributes.

A similar approach was adopted for **Exp2**. The kind of errors introduced in this experiment is shown in the X axe of Figure 18. Note that all five decision-makers involved had poor skills in multidimensional modeling.

To complete our evaluation, we collected the feedback of the decision-makers involved in the two experiments by means of a questionnaire. The participants answered 14 questions aimed at assessing their satisfaction levels with our approach. For the first 11 questions we adopted a 5-point Likert scale [74], which allows a neutral midpoint and two nuances for positive and negative answers (i.e., 'very dissatisfied', 'fairly dissatisfied', 'neither satisfied nor dissatisfied', 'fairly satisfied', and 'very satisfied'). These questions are:

1. Overall satisfaction with the methodology
2. Success in identifying errors in cube schemata
3. Success in identifying conflicts in multidimensional elements
4. Success in building group consensus
5. Better understanding of cube schemata
6. Complexity of the endorsement process
7. Satisfaction with the user interface
8. Effectiveness of facilitation
9. Understandability of the quality attributes
10. Ease of keeping up with the overall process
11. Willingness to reuse the methodology in the future

For the remaining three open questions we gave a maximum space of two sentences to provide general suggestions.

9.2. Results

We start by discussing **Exp1**. With reference to the Filter step, Figure 15 shows the attractiveness and the trustworthiness of each indicator and level in Figure 9. Note that three elements (User, Week, and Location) have been only proposed by one volunteer, five elements by all the volunteers. This confirms that, while there is some general agreement among volunteers about the elements that should be used for biodiversity analysis, some volunteers may provide particular suggestions, whose validity is to be confirmed by decision-makers. Since our experiment took place within a controlled environment, to compute trustworthiness we chose $w^{exp} = w^{rep} = 0.5$. We opted for a 100% filtering threshold, so five elements (those with 100% trustworthiness) were automatically approved, while the remaining seven had to be validated

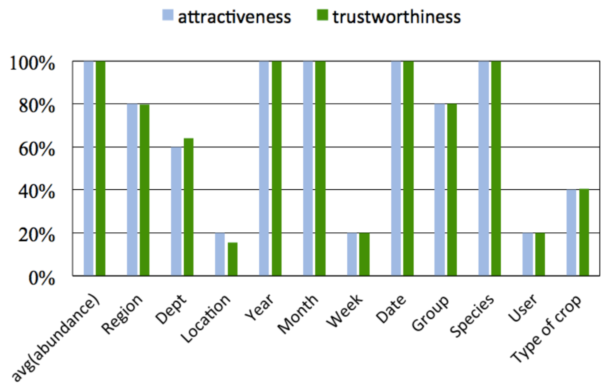

Figure 15: Attractiveness and trustworthiness of multidimensional elements (indicator and levels) in **Exp1**

by decision-makers. Eventually, only the three elements with lowest trustworthiness (namely, Location, Week, and User) were edited (Week and User were removed, Location was renamed to City), which confirms the effectiveness of this quality attribute for filtering.

The results of the *Endorse* step are shown in Figure 16. All members of group B found all problems except one (relevance of the User level), which was identified by two of the three users. The overall voting process took about 45 minutes. Conversely, for group A, one problem (completeness of the set of dimensions) was not detected by any member, while only two problems were detected by all three members. For this group, the overall endorsement process (preliminary analysis plus global discussion) took about 30 minutes.

In the *Edit* step, all the actions required to fix the problems detected were carried out. Remarkably, group A had not detected the problem related to the completeness of the set of dimensions, so (differently from group B) it did not add dimension Altitude. Besides, group A had not agreed on the minimality issue for the set of hierarchies of the Time dimension, while group B chose to delete the Time_month hierarchy.

As to **Exp2**, Figures 17 and 18 show the attractiveness and trustworthiness of multidimensional elements and the number of decision-members who individually found quality problems for the second cube schema. Also these results confirm the usefulness of the *Filtering*, *Endorse*, and *Edit* steps of our methodology.

9.3. Discussion

Overall, our two experiments show that (i) the metric measuring the trustworthiness of multidimensional elements can reliably predict their correctness, so it can be used to relieve decision-makers from the tedious task of validating all elements; (ii) the attributes measuring the quality of elements offer decision-makers a valid support in detecting possible problems with cube schemata. In

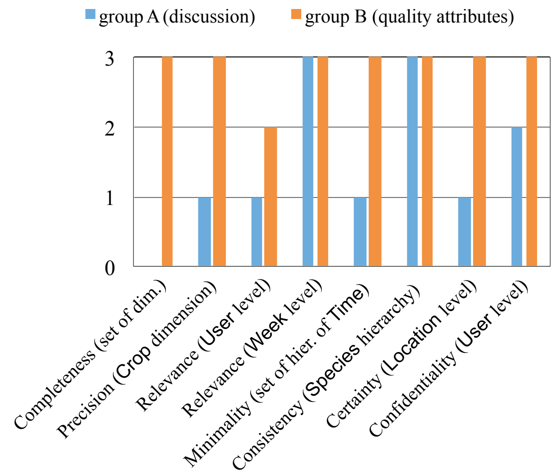

Figure 16: Number of decision-members who individually found quality problems in **Exp1**

particular, the experimental results suggest that, by relying on these attributes, a larger number of decision-makers can individually detect quality problems in the schema. This is encouraging because it suggests that adopting this methodology would allow to reduce the number of decision-makers in the *Endorse* step, leading to a significant saving in time and effort, without compromising the quality of the resulting schemata.

Clearly, the choice of the threshold for filtering has a crucial role in ensuring the success of our methodology. Indeed, the lower the threshold, the lower the number of elements to be manually evaluated. Since the duration of the Endorse step is obviously proportional to the number of multidimensional elements evaluated, the threshold should be set according to the available time of the decision-makers. For example, in **Exp1**, using a 100% threshold resulted in the GDSS session taking less than one hour. On the other hand, having some multidimensional elements automatically approved without any manual check may cause quality issues, so when decision-makers are closely involved in the project a high threshold should be preferred.

Finally, the results of the questionnaire can be judged to be satisfactory. The questions with the most critical feedback are 6, 9, and 10. This has also been orally expressed by more than 50% of the decision-makers. More precisely, the quality attributes for endorsement support were not immediately understandable, and the facilitator had to repeat their explanation at each voting step. This has made facilitation more complex. However, the general understanding of the endorsement process and of the quality attributes has shown an improvement over time. In fact, though decision-makers found it hard to keep up with the endorsement of the first two-three elements, later on the process became clearer so they could better focus on the subject matter rather than on the process itself.

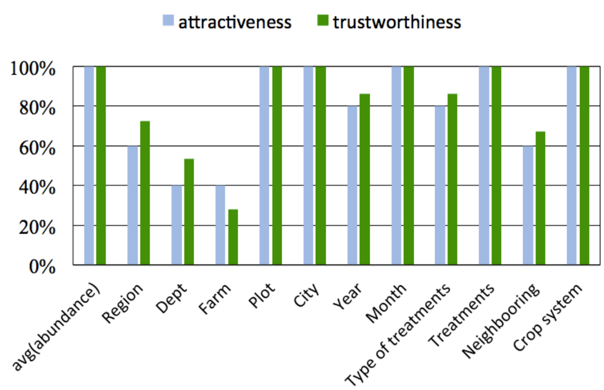

Figure 17: Attractiveness and trustworthiness of multidimensional elements (indicator and levels) in **Exp2**

10. Conclusion

The more faithfully a cube schemata represents the end-users' requirements, the more successful a DW project is. Thus, ensuring that multidimensional modeling properly takes into account all requirements is crucial. Unfortunately, classical methodological approaches fall short when a large number of end-users are involved, so collaborative approaches are necessary in this case. Achieving an effective collaboration is particularly challenging in projects where end-users are volunteers and are not really engaged (e.g., in citizen science projects), due to their varying reliability, expertise, and reputation.

To cope with these issues, in this work we have proposed a framework for multidimensional modeling that can adapt itself to the different profiles of the actors involved. To support the approach, we have proposed two sets of quality attributes related to cube schemata and end-users, respectively. Finally, we have described an implementation that relies on a web-based GDSS, and we have assessed its effectiveness using a real case study concerning biodiversity in the agricultural context.

Our future work on this topic will be to enhance the collaborative methodology by properly taking into account the discussions and comments that volunteers leave on the Wiki-based OLAP front-end used for rapid prototyping of cube schemata. To investigate the scalability of our approach, we also plan to test it on a case study involving a larger number of end-users and a more complex schemata.

Acknowledgement

This work was fully supported by the French National Research Agency as part of the the project VGI4bio (ANR-17-CE04-0012).

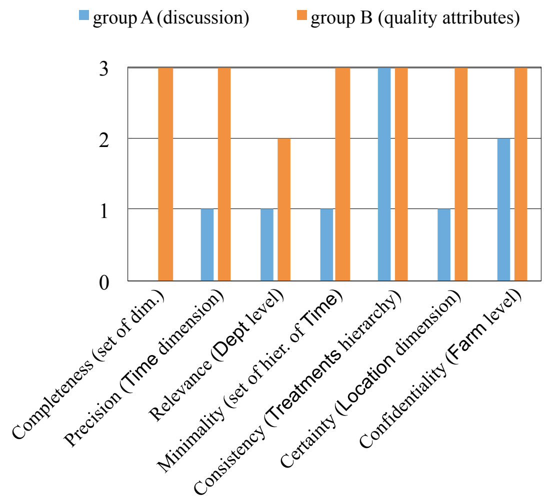

Figure 18: Number of decision-members who individually found quality problems in **Exp2**

References

[1] R. Kimball, M. Ross, The data warehouse toolkit: the complete guide to dimensional modeling, 2nd Edition, Wiley, 2002.

[2] E. Gallinucci, M. Golfarelli, S. Rizzi, Approximate OLAP of document-oriented databases: A variety-aware approach, Inf. Syst. 85 (2019) 114–130.

[3] E. Gallinucci, M. Golfarelli, S. Rizzi, A. Abelló, O. Romero, Interactive multidimensional modeling of linked data for exploratory OLAP, Inf. Syst. 77 (2018) 86–104.

[4] W. Yeoh, A. Koronios, Critical success factors for business intelligence systems, JCIS 50 (3) (2010) 23–32.

[5] M. Golfarelli, S. Rizzi, E. Turricchia, Modern software engineering methodologies meet data warehouse design: 4WD, in: Proc. DaWaK, Toulouse, France, 2011, pp. 66–79.

[6] D. C. Brabham, Crowdsourcing as a model for problem solving: An introduction and cases, Convergence 14 (1) (2008) 75–90.

[7] A. Sakka, S. Bimonte, L. Sautot, G. Camilleri, P. Zaraté, A. Besnard, A volunteer design methodology of data warehouses, in: Proc. ER, Xi'an, China, 2018, pp. 286–300.

[8] B. Gavish, J. H. Gerdes Jr., Voting mechanisms and their implications in a GDSS environment, Annals OR 71 (1997) 41–74.

[9] S. Bimonte, S. Rizzi, L. Sautot, B. Fontaine, Volunteered multidimensional design to the test: The farmland biodiversity VGI4Bio project's experiment, in: Proc. DOLAP@EDBT/ICDT, Lisbon, Portugal, 2019, pp. 1–10.

[10] A. Ballatore, Defacing the map: Cartographic vandalism in the digital commons, CoRR abs/1404.3341 (2014).

[11] C. Batini, C. Cappiello, C. Francalanci, A. Maurino, Methodologies for data quality assessment and improvement, ACM Comput. Surv. 41 (3) (2009) 16:1–16:52.

[12] R. Winter, B. Strauch, A method for demand-driven information requirements analysis in data warehousing projects, in: Proc. HICSS, 2003, pp. 1359–1365.

[13] B. Hüsemann, J. Lechtenbörger, G. Vossen, Conceptual data warehouse design, in: Proc. DMDW, 2000, pp. 3–9.

[14] M. Golfarelli, D. Maio, S. Rizzi, The Dimensional Fact Model: A conceptual model for data warehouses, IJCIS 7 (2-3) (1998) 215–247.

[15] R. Bruckner, B. List, J. Schiefer, Developing requirements for data warehouse systems with use cases, in: Proc. Americas Conference on Information Systems, 2001, pp. 329–335.

[16] J. Mazón, J. Trujillo, M. Serrano, M. Piattini, Designing data warehouses: From business requirement analysis to multidimensional modeling, in: Proc. Int. Work. on Requirements Engineering for Business Needs and IT Alignment, 2005.

[17] A. Bonifati, F. Cattaneo, S. Ceri, A. Fuggetta, S. Paraboschi, Designing data marts for data warehouses, ACM Transactions on Software Engineering Methodologies 10 (4) (2001) 452–483.

[18] P. Giorgini, S. Rizzi, M. Garzetti, GRAnD: A goal-oriented approach to requirement analysis in data warehouses, Decis. Support Syst. 45 (1) (2008) 4–21.

[19] O. Romero, A. Abelló, Multidimensional design by examples, in: A. M. Tjoa, J. Trujillo (Eds.), Proc. DaWaK, Krakow, Poland, 2006, pp. 85–94.

[20] T. Niemi, J. Nummenmaa, P. Thanisch, Constructing OLAP cubes based on queries, in: Proc. DOLAP, Atlanta, Georgia, USA, 2001, pp. 9–15.

[21] R. Nair, C. Wilson, B. Srinivasan, A conceptual query-driven design framework for data warehouse, World Academy of Science, Engineering and Technology 25 (2007) 141–146.

[22] O. Romero, A. Abelló, A survey of multidimensional modeling methodologies, IJDWM 5 (2) (2009) 1–23.

[23] N. Kozmina, L. Niedrite, OLAP personalization with user-describing profiles, in: Proc. BIR, Rostock Germany, 2010, pp. 188–202.

[24] J. Aligon, M. Golfarelli, P. Marcel, S. Rizzi, E. Turricchia, Mining preferences from OLAP query logs for proactive personalization, in: Proc. ADBIS, Vienna, Austria, 2011, pp. 84–97.

[25] S. Bimonte, L. Antonelli, S. Rizzi, Requirements-driven data warehouse design based on enhanced pivot tables, Journal of Requirement Engineering (to appear).

[26] M. L. Chouder, R. Chalal, W. Setra, Requirement analysis in data warehouses to support external information, in: Proc. CIIA, Saida, Algeria, 2015, pp. 243–253.

[27] E. Elamin, S. Alshomrani, J. Feki, SSReq: A method for designing star schemas from decisional requirements, in: Proc. ICCCCEE, 2017, pp. 1–7.

[28] J. Mazón, J. Pardillo, E. Soler, O. Glorio, J. Trujillo, Applying the i* framework to the development of data warehouses, in: Proc. iStar, Recife, Brazil, 2008, pp. 79–82.

[29] N. Ahmadi, M. Jazayeri, F. Lelli, S. Nesic, A survey of social software engineering, in: Proc. ASE, L'Aquila, Italy, 2008, pp. 1–12.

[30] I. Lykourentzou, F. Dagka, K. Papadaki, G. Lepouras, C. Vassilakis, Wikis in enterprise settings: a survey, Enterp. Inf. Syst. 6 (1) (2012) 1–53.

[31] M. Paasivaara, S. Durasiewicz, C. Lassenius, Using scrum in a globally distributed project: a case study, Softw. Process. Improv. Pract. 13 (6) (2008) 527–544.

[32] M. A. Jeusfeld, C. Quix, M. Jarke, Design and analysis of quality information for data warehouses, in: Proc. ER, Singapore, 1998, pp. 349–362.

[33] M. A. Serrano, J. Trujillo, C. Calero, M. Piattini, Metrics for data warehouse conceptual models understandability, Inf. Softw. Technol. 49 (8) (2007) 851–870.

[34] N. Prat, S. S. Cherfi, Multidimensional schemas quality assessment, in: Proc. CAiSE Workshops, Klagenfurt/Velden, Austria, 2003, pp. 341–352.

[35] C. Calero, M. Piattini, C. Pascual, M. A. Serrano, Towards data warehouse quality metrics, in: Proc. DMDW, Interlaken, Switzerland, 2001, pp. 2.1–2.10.

[36] G. Papastefanatos, P. Vassiliadis, A. Simitsis, Y. Vassiliou, Design metrics for data warehouse evolution, in: Proc. ER, 2008, pp. 440–454.

[37] M. Golfarelli, S. Rizzi, Data warehouse testing: A prototype-based methodology, Inf. Softw. Technol. 53 (11) (2011) 1183–1198.

[38] H. Lenz, A. Shoshani, Summarizability in OLAP and statistical data bases, in: Y. E. Ioannidis, D. M. Hansen (Eds.), Proc. SSDBM, Olympia, Washington, USA, 1997, pp. 132–143.

[39] W. Lehner, J. Albrecht, H. Wedekind, Normal forms for multidimensional databases, in: M. Rafanelli, M. Jarke (Eds.), Proc. SSDBM, Capri, Italy, 1998, pp. 63–72.

[40] Y. Wand, R. Y. Wang, Anchoring data quality dimensions in ontological foundations, Commun. ACM 39 (11) (1996) 86–95.

[41] M. Allahbakhsh, A. Ignjatovic, B. Benatallah, S. Beheshti, E. Bertino, N. Foo, Reputation management in crowdsourcing systems, in: Proc. CollaborateCom, Pittsburgh, PA, USA, 2012, pp. 664–671.

[42] Y. Zhang, M. van der Schaar, Reputation-based incentive protocols in crowdsourcing applications, in: Proc. INFOCOM, Orlando, FL, USA, 2012, pp. 2140–2148.

[43] C. Barron, P. Neis, A. Zipf, A comprehensive framework for intrinsic OpenStreetMap quality analysis, Trans. GIS 18 (6) (2014) 877–895.

[44] P. Mooney, P. Corcoran, How social is OpenStreetMap, in: Proc. GIScience, Avignon, France, 2012, pp. 24—27.

[45] M. Napolitano, P. Mooney, MVP OSM: a tool to identify areas of high quality contributor activity in OpenStreetMap, The Bulletin of the Society of Cartographers 45 (1) (2012).

[46] P. Neis, A. Zipf, Analyzing the contributor activity of a volunteered geographic information project — the case of OpenStreetMap, ISPRS Int. J. Geo-Inf 1 (2) (2012).

[47] S. Scheider, C. Keßler, J. Ortmann, A. Devaraju, J. Trame, T. Kauppinen, W. Kuhn, Semantic referencing of geosensor data and volunteered geographic information, in: N. Ashish, A. P. Sheth (Eds.), Geospatial Semantics and the Semantic Web - Foundations, Algorithms, and Applications, Springer, 2011, pp. 27–59.

[48] P. Neis, M. Goetz, A. Zipf, Towards automatic vandalism detection in OpenStreetMap, ISPRS Int. J. Geo-Information 1 (3) (2012) 315–332.

[49] F. D'Antonio, F. Paolo, T. Kauppinen, VGI edit history reveals data trustworthiness and user reputation, in: Proc. AGILE Conf. on Geographic Information Science, Castellon De La Plana, Spain, 2014, pp. 1–5.

[50] A. Alghanim, Assessing OpenStreetMap road semantics quality using context, data trustworthiness and user reputation, Ph.D. thesis, University College, Dublin, Ireland (2020).

[51] T. D. Huynh, N. R. Jennings, N. R. Shadbolt, An integrated trust and reputation model for open multi-agent systems, Auton. Agents Multi Agent Syst. 13 (2) (2006) 119–154.

[52] J. F. N. Jr., R. O. Briggs, D. D. Mittleman, D. R. Vogel, P. A. Balthazard, Lessons from a dozen years of group support systems research: A discussion of lab and field findings, J. of Management Information Systems 13 (3) (1997) 163–207.

[53] R. O. Briggs, G.-J. D. Vreede, J. F. N. Jr, Collaboration engineering with ThinkLets to pursue sustained success with group support systems, J. of Management Information Systems 19 (4) (2003) 31–64.

[54] A. Fernandez, E. Insfrán, S. Abrahão, Usability evaluation methods for the web: A systematic mapping study, Inf. Softw. Technol. 53 (8) (2011) 789–817.

[55] E. Carmel, R. D. Whitaker, J. F. George, PD and joint application design: A transatlantic comparison, Commun. ACM 36 (6) (1993) 40–48.

[56] C. McGoff, A. Hunt, D. Vogel, J. Nunamaker, IBM's experiences with GroupSystems, Interfaces 20 (6) (1990) 39–52.

[57] R. Evans, S. Park, H. Alberts, Decisions not requirements: decision-centered engineering of computer-based systems, in: Proc. ECBS, Monterey, CA, USA, 1997, pp. 435–442.

[58] G. Ruhe, Software engineering decision support: Methodology and applications, in: Innovations in Decision Support Systems, Vol. 3 of International Series on Advanced Intelligence, 2003, pp. 143–174.

[59] A. Aurum, C. Wohlin, The fundamental nature of requirements engineering activities as a decision-making process, Inf. Softw. Technol. 45 (2003) 945–954.

[60] G. de Vreede, R. O. Briggs, A program of collaboration engineering research and practice: Contributions, insights, and future directions, J. of Management Information Systems 36 (1) (2019) 74–119.

[61] F. Ricci, L. Rokach, B. Shapira (Eds.), Recommender Systems Handbook, Springer, 2015.

[62] A. Gunawardana, G. Shani, Evaluating recommender systems, in: F. Ricci, L. Rokach, B. Shapira (Eds.), Recommender Systems Handbook, Springer, 2015, pp. 265–308.

[63] R. K. Wong, V. W. Chu, T. Hao, J. Wang, Context-aware service recommendation for moving connected devices, in: Proc. ICCVE, Beijing, China, 2012, pp. 136–141.

[64] K. Boulil, S. Bimonte, F. Pinet, Conceptual model for spatial data cubes: A UML profile and its automatic implementation, Comput. Stand. Interfaces 38 (2015) 113–132.

[65] S. Bimonte, É. Edoh-Alove, H. Nazih, M. Kang, S. Rizzi, ProtOLAP: rapid OLAP prototyping with on-demand data supply, in: I. Song, L. Bellatreche, A. Cuzzocrea (Eds.), Proc. DOLAP, San Francisco, CA, USA, 2013, pp. 61–66.

[66] M. Banek, B. Vrdoljak, A. M. Tjoa, Z. Skocir, Automating the schema matching process for heterogeneous data warehouses, in: Proc. DaWaK, Regensburg, Germany, 2007, pp. 45–54.

[67] R. Torlone, Two approaches to the integration of heterogeneous data warehouses, Distributed Parallel Databases 23 (1) (2008) 69–97.

[68] E. Rahm, P. A. Bernstein, A survey of approaches to automatic schema matching, VLDB J. 10 (4) (2001) 334–350.

[69] E. S. Raymond, The cathedral and the bazaar - musings on Linux and Open Source by an accidental revolutionary, O'Reilly, 1999.

[70] G. Poels, G. Dedene, DISTANCE: a framework for software measure construction, Tech. rep., Departement Toegepaste Economische Wetenschappen (1999).

[71] G. Poels, On the formal aspects of the measurement of object-oriented software specifications, Ph.D. thesis, aculty of Economics and Business Administration, Katholieke Universiteit, Leuven, Belgium (1999).

[72] G. L. Kolfschoten, R. O. Briggs, G. de Vreede, P. H. M. Jacobs, J. H. Appelman, A conceptual foundation of the thinklet concept for collaboration engineering, Int. J. Hum. Comput. Stud. 64 (7) (2006) 611–621.

[73] O. Billaud, R.-L. Vermeersch, E. Porcher, Citizen science involving farmers as a means to document temporal trends in farmland biodiversity and relate them to agricultural practices, Journal of Applied Ecology, to appear (2020).

[74] J. T. Nadler, R. Weston, E. C. Voyles, Stuck in the middle: The use and interpretation of mid-points in items on questionnaires, The J. of General Psychology 142 (2) (2015) 71–89.