Context in Spreadsheet Comprehension

Andrea Kohlhase
Information Management
University of Applied Sciences Neu-Ulm
Germany

Michael Kohlhase
Computer Science
Jacobs University Bremen
Germany

Ana Guseva
Computer Science
Macedonia

***Abstract*—Even though spreadsheet programs traditionally concentrate on exploration and computation of data (the author's view), a non-trivial proportion of spreadsheets are used for communicating data, models, and decisions to humans who assume the role of a spreadsheet reader. The communicative use of spreadsheets gives the spreadsheet context, that is, the background knowledge needed to interpret and make sense of its content, a very important role. Indeed, many 'spreadsheet errors' can be traced to mis-interpretations due to context failures.**

In this paper we report on a set of experiments we conducted to get a deeper understanding of the context in spreadsheet comprehension, focusing on the different perspectives authors and readers take in understanding spreadsheets. The results confirm missing context information as a likely source for semantic spreadsheet errors. Moreover, they lead to an extension and refinement of already established context dimensions.

I. Introduction

In terms of adoption spreadsheet software is extremely successful. Spreadsheet documents have developed from being easy-to-use, programmable interfaces into easy-to-understand, sharable data interfaces (see e.g., [1], [2]).

Hence, the user experience does not stop with the creation of a spreadsheet document, it also involves others reading it later. The infamous high error-rate in spreadsheet documents (see [3], [4], [5]) invoked an intense rush to studies and solutions over the last couple of years, most of which tend to address the avoidance of introducing errors when authoring a spreadsheet. Our scientific interest revolves around the spreadsheet document as a medium of communication, therefore we are particularly interested in the reader's experience of spreadsheets as a data interface and concerned about the errors introduced in the comprehension process.

A taxonomy of the errors [6] shows that a significant portion of errors (87%, as calculated in [4]) are semantic. In this research, a **semantic error** is one that is committed when users have a wrong concept that may be correctly or incorrectly put into practice. These arise from misunderstanding the real-world, wrong translation of the real-world to the spreadsheet representation, or a misunderstanding of the spreadsheet's internal logic [4]. As semantic errors are made on an individual document base, there is neither hope for a best-practice guide to train avoiding them nor for a general software update to help out. Semantic errors pose a more serious threat for wide-impact spreadsheets since more and more individual communication errors might aggregate over the span of distribution.

It has been proposed [4], [7] that a key reason in committing semantic errors is a missing higher-level abstraction of the data. Tables, with their grid framework, expose details and allow manipulation of underlying data. Therefore, spreadsheets, as a computer-supported realization of tables, turn one's attention to data on a micro-level, failing to provide the big picture. Generally, schematic diagrams or pictures abstract away and integrate the data, presenting it holistically. Newer versions of spreadsheet software like "MS Excel 2013" thus have powered up their visualization features, e.g. with "Power View". In particular, Power View provides users with report features and analytical views, which helps them to comprehend the spreadsheet data and, thus, in the end increase the spreadsheet's readability.

The communicative aspect of spreadsheets allows us to understand them as knowledge sharing tool. Osterlund and Carlile describe the semantic issues with knowledge sharing as follows:

> "*the relational core of a knowledge sharing theory easily falters. [...] We end up instead with a perspective that focuses on the storage and retrieval of explicit knowledge represented in information systems. Knowledge becomes an object shared within and across community boundaries without consequence for the community in which it originated*". [8, p. 18]

Note that crossing a community boundary leaves the entire **context** – the circumstances and settings in which a document is created and obtains its specific meaning – behind. Researchers in the field of Human-Computer Interaction (HCI) have focused in recent years on the context-of-use of software systems: user experience issues often only arise in the concrete context in which a product is used. Our approach for tackling the readability issue of spreadsheets is motivated by this insight. Therefore we ask: what is the context of a spreadsheet document and which role does it play for comprehension of spreadsheets? For an answer, consider the following distinct contexts:

- the context of the data itself,
- the information context (implicit knowledge) of the author or the reader,
- the event context of the author (the intention of the document as communication tool) or the reader (the expectation towards the usefulness of the document),
- the effect context (e.g. decision making based on the document).

Note that the clear distinction between authors and readers is

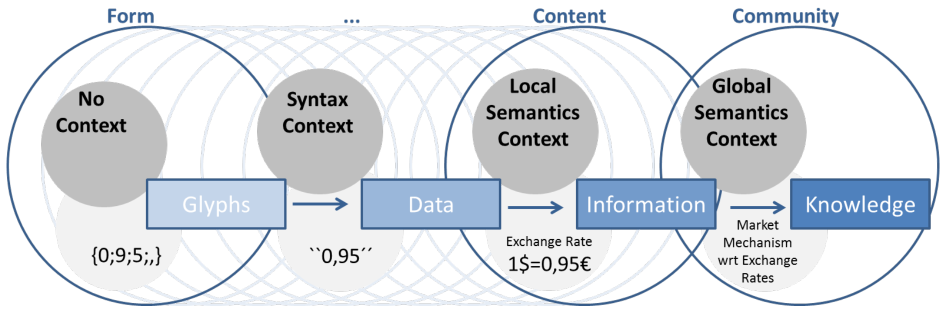


Fig. 1: From Mere Glyphs To Valuable Knowledge

only an analytical one. We are well aware that authors turn into readers after a short while even for their own documents and, vice versa, that the motivation of readers might consist in searching for copy-able parts to author their very own spreadsheets. Nevertheless, the context can be clearly distinguished where wide-impact, local boundary-leaving spreadsheets are concerned.

Probst et al. (see [9]) posit that glyphs, data, information, and knowledge can be seen as stages of a pipeline (shown in Fig. 1) being defined by the respective context. In particular, they argue that *glyphs* are just a set of characters or symbols like {0;9;5;,} without any structure. A first set of rules imposed on the glyphs — the syntax context — then yields *data* which can be handled by machines. Spreadsheet values are data and as such they can be computed by the calculation engine of a spreadsheet software. For obtaining meaning from such data we still need another component: the context of meaning. Usually, we discern data from *information* by viewing information as data with a meaning. Davenport and Prusak think of information "*as data that makes a difference*" [10]. Data becomes information if a user can interpret the data in regard to a specific goal, that is a local meaningful context like using the string '0,95' as number in an equation concerning exchange rates in our example. In contrast, information becomes *knowledge*, if a user can interpret the information in regard to a global context of meaning like understanding the exchange rate equation in the area of specific market behavior with respect to change of exchange rates. Therefore, the role of context for a spreadsheet consists in turning mere values into content. We can even say that we get the more content the more context there is. So far, the context which allows spreadsheets to communicate information is contained in the table, row and column headers and sometimes in added comments.

Unfortunately, more often than not a spreadsheet can neither be properly understood nor used, unless one looks beyond the spreadsheet data itself to the broader background knowledge within which it is embedded (see for instance [11]). This contextual knowledge is thus a frame and provides means for the spreadsheet's appropriate interpretation.

In this paper we are aiming at a better understanding of what spreadsheet context is, refining and extending a previous study presented in [7]. Here, we compare spreadsheet authors' understanding of context to spreadsheet readers'; they turn out to be very different. As the complexity of a spreadsheet document might be a relevant distinction with regard to context, we presented readers and authors with a simple and a complex spreadsheet. In Section II we survey previous research concerning semantic spreadsheet errors, in Section III we present the details, results and interpretation of our study and we conclude in Section IV.

II. State of the Art

There is on-going research [12], [7], [13] on ways to reduce spreadsheet errors. Some approaches [3], [6] focus on identifying best practices for error prevention, while others [12], [13] tackle the root problem and concentrate on empowering spreadsheets with better ways of interaction.

Regarding error prevention, the approach is in understanding common errors, so that they may be avoided. Much research has shown that errors may not be eliminated completely, but rather, reduced [3]. Best practices for this are often proposed based on good software development principles. For instance, one suggestion is having a full system development life cycle with requirements analysis and design stages (both of which are largely skipped). Other tactics, such as layout-planning, cell-protection, even modular design, have been suggested to protect against some errors, yet, all of these do not address another problem - that of the semantic errors committed.

Nardi and Miller noticed a relevant feature of spreadsheets in [1]: they are not "single-user applications". In particular, they are used in the work environment as collaboration tool and as means of communication to exchange and combine domain knowledge and programming expertise. Even though "*the visual clarity of the spreadsheet table exposes the structure and intent of users' models, encouraging the sharing of domain knowledge*" [14], this doesn't mean that there is no information loss in the sharing process. As Hendry and Green in [15] point out, "*the main resources available to spreadsheet users to improve comprehensibility are to use titles and to arrange the layout carefully*". These resources may be sufficient for simple spreadsheets, but for complex ones they fail as the high semantic error rate in [4] indicates.

To address this problem, Green et al. argued in [16] that a solution might be a "*browsable description level*" to provide more context. Essentially, a scheme for attaching descriptions in which attributes and relationships can be recorded and later searched for. It is effective in understanding and reusing code, which in this case is a large and complex body of information. Hendry and Green [15] have imported this idea to spreadsheet design. "CogMap", a user assistance tool developed for this purpose, provides capabilities for simple annotation of spreadsheets with 'tags', which can later be filtered in color-coded views. Tagging and annotating regions of spreadsheet data provide simple, off-hand taxonomies, but cannot describe richer information structures as ontologies can.

In fact, by now ontologies have found their way in many spreadsheet user assistance tools. For example, "RightField" [17] is an open source application that provides a mechanism for embedding ontology annotation support in Excel spreadsheets. Another such tool is our own "SACHS" system [12] with which ontology-based enhancement of some context aspects can be developed. A reader, in turn, can be informed about these aspects later on: the system allows the author to state his/her domain knowledge in a structured ontology which SACHS integrates into the spreadsheet in form of adapted help texts. [7] shows that authors describing implicit knowledge, use a mix of the following context dimensions:

- Definition (conceptual) – a thorough description of the meaning
- Purpose (conceptual) – the intention, i.e. why a specific information is put there
- Assessment of Purpose – an interpretation of the purpose so that it allows drawing conclusions/actions
- Assessment of Value – an interpretation of data so that it allows for making judgement e.g. "if the ratio is close to 100% everything is fine"
- Formula – description of data by specifying how it was computed (what function)
- Provenance – the source of the data i.e. how it was obtained (direct measurement, computation, import etc.).
- History – explanation of how the spreadsheet was changed over time

On the other end of the communication pipeline stand the spreadsheet readers. As we observe in [18] research so far has accented usability problems for developers of spreadsheets, whereas the readers are largely overlooked. [18] tries to remedy this and studies how readers perceive information offered in spreadsheets. The spreadsheet authors' context dimensions in [7] and the readers' information models in [18] are the starting point of this research.

III. The Study

The study we report on is based on a Bachelor Thesis project conducted by Ana Guseva in [19] under the other authors' close supervision. We use the data collected there, but refine the interpretation in this paper.

We conducted 6 interviews with three participants representing each target group:

- **Authors** who have developed the particular or a similar spreadsheet, thus are experts in the spreadsheet knowledge domain, and
- **Readers** who are only moderately familiar with the spreadsheet knowledge domain, but are directly affected by the spreadsheet and therefore motivated to read it.

All participants were technologically fluent in computer use and had at least basic experience with spreadsheet technology. The interviews were then transcribed into 27 pages of interview material.

Each interview covered 2 test cases each of which centered around a distinct spreadsheet. These spreadsheets were developed in 2010 for budget planning in a research project and are currently still in use:

- **Complex** As a complex spreadsheet we used a cluttered spreadsheet that contains
  - numerous data spread over multiple screens,
  - complex formulae for calculations making use of the more advanced functions offered by the computation engine, and
  - multiple nested data dependencies;
- **Simple** Our simple spreadsheet is
  - neatly organized on one screen,
  - uses only basic formulae drawing on basic functions like `sum` and `divide`, and
  - containing only direct and somewhat expected data dependencies like the ones in a total sum.

The choice of these different kinds of spreadsheet was intended to investigate the influence of spreadsheet complexity on the presence and relevance of the specific context. The distinctions allow us to judge how successfully the authors' context knowledge is transferred to the readers via the spreadsheet document. The data collection procedures involved tape recorded interviews. The interviews were conversational in style, they intended to capture the users' understandings in their own words and communication style. A fixed set of open-ended questions was presented to each user, yet the questions were asked as they arose naturally in the context of the conversation.

In particular, the exact interview procedure used is a variant of the **Wizard of Oz (WOZ)** technique [20], in which participants are given the impression that they are interacting with a program, when in fact the program is operated by an invisible human – the wizard. It is popular in the fields of experimental psychology, human factors, ergonomics, linguistics, and interface and usability engineering [21].

The approach taken extends the application of the WOZ technique, by reversing it. Previously used in [7], the **Reverse Wizard of Oz** experiment, puts the participant in the position of an ideal spreadsheet system, while the investigator interacts with it, asking for help. It enables the investigator to act as if she is unfamiliar with the test spreadsheet, and thus, allows her to ask the participant to thoroughly and in detail explain the spreadsheet. This in turn, elicits the participant's contextual knowledge.

A. Data Analysis Method: Card Sorting

The next step in the investigation procedure was the **card sorting**. As [22] puts it, "*Card sorting is a technique to understand relationships between items, to group items into dimensions, and to understand users' mental models of item organization*".

Traditionally, card sorting is used in designing information architecture, workflows, menu structure and web-site navigation paths. It involves many participants sorting a given set of cards according to their best understanding. In this research only the investigator sorted the cards. Effectively, here the card sorting reflects the investigator's mental model of what the participants had in mind.

To obtain cards the interview data was split up into units called **knowledge items**, i.e., the smallest, still meaningful parts of sentences extracted from the transcribed interview. Here are some examples:

- "*All the costs [G11-23] are summed up in [G24]*".
- "*I think that's a summary table of the above.*"
- "*You can do it and insert another column (if you need to add more partners).*"

Note that each knowledge item carries one complete piece of information, so we created a card for each knowledge item. A total of 319 cards were generated from the interview material.

As is the practice in open coding [23], [21], [24], the cards were compared with others for similarities and differences. In this way, conceptually similar cards were grouped together to form dimensions and subdimensions. Making use of constant comparisons guarded against the researcher's bias and achieved both greater precision (the grouping of like and only like phenomena) and consistency (always grouping like with like) [23]. The dimensions[1] were then given labels, depending on the conceptual notion they denoted. Once identified, dimensions and their properties became the basis for sorting the next cards. In sorting the next card, notice was taken of similar cards already present in other dimensions. In fact, during this process comparative questions were used, such as: 1) What is the essence of this card?, 2) How does it compare to the previous cards?, 3) How does it differ from the previous cards?, 4) Does it fit in any previous dimension?

It must be noted, that a single card, if necessary, was placed under one or more dimensions. Additionally, if during the sorting, a previously sorted card was noticeably no longer fit for a dimension, it was moved to another dimension where it belonged. An additional closed card sorting (which verifies dimensions) followed the open card sorting (which creates new dimensions) to further consolidate the results.

B. Data Analysis: Context Dimensions

The intermediate result of a card sorting process consists of a set of card piles. These piles represent the distinct dimensions of the context space of spreadsheets. Here, all 319 cards (knowledge items) were sorted into 14 piles.

[1]Note that we use the term "dimension" to be consistent with [7].

In the next step each pile is labeled, so that we get a set of names for distinct context dimensions. In Table I these are shown together with the defining question(s) for each respectively.

Dimension	Question
STATEMENT	*What is it? (read keyword)*
REPHRASING	*What is it? (rephrase keyword)*
DEFINITION	*What is it? (formal definition)*
BY-EXAMPLE	*Example?*
EVALUATION	*Is it good?*
FORMULA	*How is it calculated? (function)*
PROVENANCE	*From where?*
	How is it calculated? (dependency)
REASON	*How come?*
	Why is it so?
HISTORY	*Has it changed?*
ORGANIZATION	*Where is it?*
USAGE	*What should we do with it?*
PURPOSE	*For what do we need it?*
SIGNIFICANCE	*Is it important?*
OTHER	n/a

TABLE I: Dimensions and Corresponding Questions

The dimensions in grey were so small, that we dropped them in the following to prevent confusion.

In Table II we listed the occurrence rate $c_{\text{DIMENSION}}/c$ of cards belonging to a respective context dimension differentiating by user role and the complexity level of the spreadsheet, where $c_{\text{DIMENSION}}$ = number of cards in resp. context dimension and c = total number of cards. Note that some cards were sorted into multiple dimensions, so that the percentages don't add up to 100%.

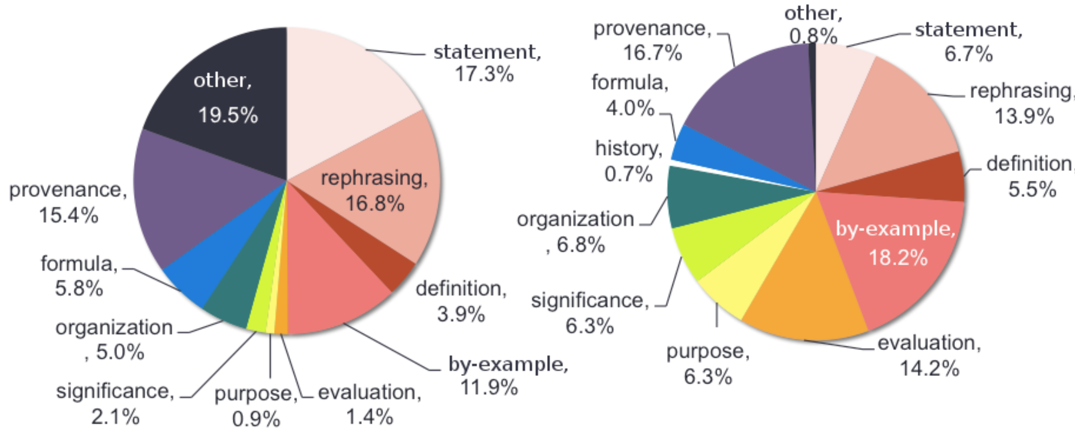
Dimension	Complex		Simple	
	Authors	Readers	Authors	Readers
STATEMENT	13.8%	27.6%	14.1%	30.3%
REPHRASING	35.9%	25.9%	22.5%	30.3%
DEFINITION	13.1%	6.9%	9.9%	6.1%
BY-EXAMPLE	36.6%	15.5%	39.4%	24.2%
EVALUATION	28.3%	1.7%	31.0%	3.0%
FORMULA	8.3%	10.3%	8.5%	9.1%
PROVENANCE	29.0%	24.1%	40.8%	27.3%
HISTORY	2.8%	0.0%	0.0%	0.0%
ORGANIZATION	20.0%	13.8%	8.5%	3.0%
PURPOSE	15.2%	0.0%	11.3%	3.0%
SIGNIFICANCE	13.8%	6.9%	12.7%	0.0%
OTHER	0.7%	37.9%	2.8%	27.3%

TABLE II: Relevance of Dimensions given by Occurrence Rates $c_{\text{DIMENSION}}/c$

C. Data Interpretation: Context for Comprehension

In this section we interpret the context dimensions found in the card sorting with respect to their occurrence rate for authors vs. readers (see Fig. 2) or relative to the complexity of the spreadsheet (see Fig. 3). Additionally, for each dimension we provide an exemplary knowledge item.

1) STATEMENT*:* This category contains knowledge items that repeated the facts contained in the spreadsheet that were obvious to the user, for example,



(a) Readers (b) Authors

Fig. 2: Context Dimensions wrt. Spreadsheet User Roles

"*In this project* 90% *of the workload is for RTD.*"

where the according column header read "RTD"[2] and row header "Project".

STATEMENT is used by readers twice as often than by authors independently of the complexity of the spreadsheet. This is not surprising as the spreadsheet in question was unfamiliar to the readers, whereas authors still remembered at least part of it and so they did not want to state the obvious. Note that such a statement can be a mere reading aloud process without any comprehension of what is communicated.

2) REPHRASING: All knowledge items in this dimension are stating in their own words the facts contained in the spreadsheet that were obvious to the user. This kind of action presumes that the respective fact has been understood. An example is the following, where the first part is a STATEMENT and the second part a REPHRASING:

"*So* 90 *is the total person-months for Jacobs, it is what the EU would have paid us if the project had been granted.*"

Interestingly, REPHRASING was used more by authors than readers for the complex spreadsheet (35.9% vs. 25.9%), whereas it was the other way round for the simple one (22.5% vs. 30.3%). We suspect that authors thought REPHRASING to be an added-value for complex content, whereas the simple content doesn't require this explanation type – being so simple. The readers viewed it differently, choosing REPHRASING as a valuable explanation type. Here, we observe that the authors' and the readers' contexts differ and thus their assessments do as well.

3) DEFINITION: This dimension contains all knowledge items that specify or define terminology used in the spreadsheet headers, e.g.,

"*FET is about Future Emergent Technologies, so some visions about future emergent technologies.*"

[2]In EU projects "RTD" abbreviates "Research and Technological Development".

DEFINITIONs were rather rarely given (max = 13.1%), but if so, then more by the authors than by the readers – especially for the complex spreadsheet. This indicates that this context dimension also belongs to the ones that would assist readers when interpreting a spreadsheet. Moreover, the inclusion of definitions into a spreadsheet might also be appreciated as added-value even by authors.

4) BY-EXAMPLE: Here, the set of knowledge items that included examples of specific concepts contained in the given spreadsheets, for instance,

"*Demonstration is, for example, to go on fairs.*"

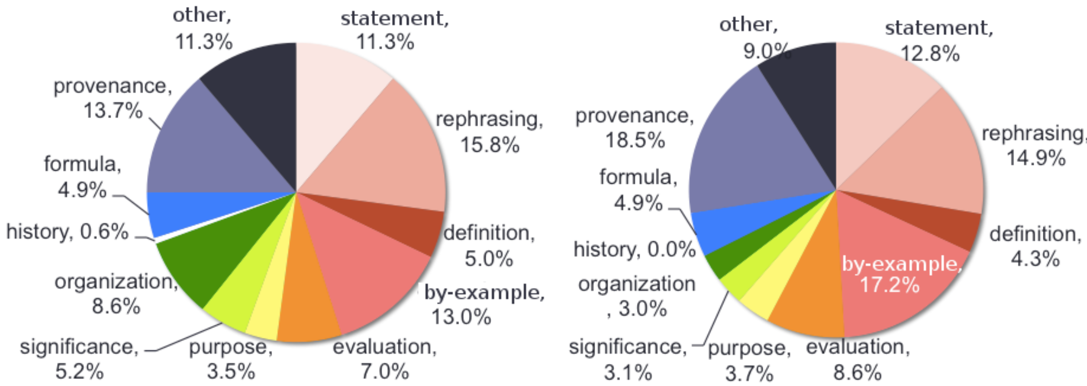
Authors were significantly more able to provide examples than readers. We can derive that examples constitute another context dimension that is missing in spreadsheets. Another interesting point wrt. BY-EXAMPLE is that authors used examples to elaborate the content almost as often for the complex as for the simple spreadsheet. We suspect therefore that they deem examples as an adequate means for explaining spreadsheets in general.

5) EVALUATION: All knowledge items that contained judgements were collected into this context dimension. Users drew consequences based on their interpretation of the spreadsheet content – for example, they assessed values or did plausibility checks on numbers or said

"*[P21] - if that's high then we are getting a lot!*"

It helped them to make sense of the data, particularly for decision-making based on this.

Readers most rarely *assessed* values (1.7% vs. 3.0%), while authors did so quite often (28.3% vs. 31.0%). This strongly demos that readers need help when interacting with spreadsheets, since they are often unable to make decisions based on the given values. As we believe that enabling decision-making is often a motivation to distribute spreadsheets, the addition of support for this context dimension would be a highly appreciated assistance in spreadsheet use.



(a) Complex (b) Simple

Fig. 3: Context Dimensions wrt. Spreadsheet Complexity

6) FORMULA: This context dimension encompasses all explanations concerning the computational aspects of formulae, for example,

> "*It [N22] is just multiplied by the person-months, but again it's the same formula as [L22].*"

The precise formula to calculate a value was relatively infrequently selected to explain spreadsheet content (max $=$ 10.3%). This is astonishing as it is supposed to be the most prominent feature of spreadsheets. We suspect that users consider this a service feature of the spreadsheet software which doesn't carry much weight in terms of comprehending a spreadsheet.

7) PROVENANCE: Knowledge items were sorted into this context dimension, if they referred to the origin of the spreadsheet content encompassing cell dependencies in formulae, for instance,

> "*It is EU guidelines with their cost model, they have invented for the different entities ....*"

It is striking that authors and readers did select PROVENANCE as context dimension to explain the complex spreadsheet much alike. On the one hand both kinds of users are able to explain the provenance. On the other hand, it is worth noting that they noticed the information about the data's provenance either present in the spreadsheet or existing in shared background knowledge. Obviously, a formula is relevant for the computation of a certain value (see context dimension FORMULA). But the presentation of the formula in a spreadsheet is also relevant for understanding the coherence of the data. In particular, a print-out of a spreadsheet document (i.e., a ledger sheet) is much less valuable as it loses provenance context. Authors especially value this information facet in formulae.

8) HISTORY: Knowledge items addressing the historical context of a spreadsheet, i.e., reflections about the creation process of a spreadsheet over time, are gathered here as, for example, this one:

> "*(It's) just some formulas ... it was inserted afterwards ... later in the table ...*"

Since spreadsheets are communication tools, they are often updated and modified by different users, which may result in changes in layout or data – sometimes leaving inconsistent, or even, superfluous information.

Naturally, this information is not reported in the spreadsheet itself and is seldom of big interest. Our spreadsheets didn't have such a long history, so that the score with respect to HISTORY was really low (max $= 2.8\%$).

9) ORGANIZATION: Here, we collected all knowledge items that concern superficial qualities of the spreadsheet, such as layout, data-format and data arrangement. In the following you find a concrete example:

> "*So the spreadsheet starts with the lower table with all the details, and the upper table is main summary, the key figures, and the table on the right is a transition especially made for the EU.*"

This dimension effects the navigation and organizational usability of the spreadsheet, for example giving different colors to those cells where the input is expected (green) or red to those cells that are out of the limit.

Authors note and point out more organizational cues than readers do. As the authors are the creators of the spreadsheet's organization they might want to point out the fine details of their design. Another interesting fact consists in the much higher rate of explanations in the ORGANIZATION context dimension by authors and readers alike for the complex spreadsheet. This means that this kind of organization also represents a valuable context dimension if the content is difficult to understand at once.

10) PURPOSE: We sorted all knowledge items that included (correct) intentional aspects of spreadsheet content into this context dimension, e.g.

> "*For each partner we need to set up a budget.*"

The PURPOSE context is implicitly hidden or explicitly missing in the spreadsheets as readers do not select this kind of explanation for the complex spreadsheet and only seldom for the simple one (max $= 3.0\%$).

11) Significance: Here, knowledge items were gathered that bear information about what the significant aspects of the spreadsheet data are, for example,

"*It's just a remark, not to forget ...*"

Significance was much more pronounced with the authors than with the readers. This is probably due to the fact that readers rarely could discern important aspects of the spreadsheet content. Significance was as important for authors for the complex as it was for the simple spreadsheet, indicating that they thought of the spreadsheet as a communication medium with a message when designing it. Readers didn't notice at all differences in relevance for the simple spreadsheet.

12) Other: This final context dimension contains all the bits and pieces that couldn't be sorted further. Additionally, it also includes utterances of wrong assumptions and definitions, doubts, guesses, and indications of ignorance like

"*'Special Transition Flat Rate' - no idea what it is!*"

This card pile really shouldn't be considered a context dimension, but it shows that readers even with a simple spreadsheet frequently make semantic errors .

D. Refinement of Context Dimensions

In Table III we compare the context dimensions obtained from the card sorting exercise reported on above, with those from [7]. Our new dimensions constitute an extension and refinement: The Definition can be refined to Statement, Rephrasing, and Definition. Similarly we now differentiate Assessment of Value into By-Example and Evaluation. The Other pile cannot be considered a context dimension, therefore we cannot consider this an extension.

Context Dimensions	in [7]
Statement	Definition
Rephrasing	Definition
Definition	Definition
By-Example	Assessment of Value
Evaluation	Assessment of Value
Significance	Assessment of Purpose
Purpose	Purpose
Organization	Purpose, Provenance
Provenance	Provenance
Formula	Formula
History	History
Other	n/a

TABLE III: Correspondence with Context Dimensions in [7]

IV. Conclusion

In this paper we have reported on a study we conducted to deepen our understanding of spreadsheet context. By interviewing authors and readers of spreadsheets and differentiating between a simple and a complex spreadsheet, we could observe clear differences between these two user roles. In general, the readers missed out on a lot of context dimensions, therefore making the case for assistance systems for spreadsheet comprehension. Moreover, we could refine the set of context dimensions given in [7].

Acknowledgment

This work has partially been supported by the German Research Council under Grant KO 2428/13-1.

References

[1] B. A. Nardi and J. R. Miller, "An ethnographic study of distributed problem solving in spreadsheet development," in *Proceedings of the 1990 ACM conference on Computer-supported cooperative work*. ACM Press, 1990, pp. 197–208.

[2] T. R. G. Green and M. Petre, "Usability analysis of visual programming environments: a 'cognitive dimensions' framework," *JOURNAL OF VISUAL LANGUAGES AND COMPUTING*, vol. 7, pp. 131–174, 1996.

[3] R. R. Panko, "What we know about spreadsheet errors," *Journal of Organizational and End User Computing (JOEUC)*, vol. 10, no. 2, pp. 15–21, 1998.

[4] M. Clermont, "A scalable approach to spreadsheet visualization," *Klagenfurt University*, 2003.

[5] K. M. Consulting, "Executive summary: Financial model review survey," KPMG, London, Tech. Rep., 1997.

[6] K. Rajalingham, D. R. Chadwick, and B. Knight, "Classification of spreadsheet errors," *arXiv preprint arXiv:0805.4224*, 2008.

[7] A. Kohlhase and M. Kohlhase, "Semantic transformation of spreadsheets," *Electronic Communications of the EASST*, vol. X, 2010. [Online]. Available: http://www.easst.org/eceasst/

[8] C. Osterlund and P. Carlile, "How practice matters: A relational view of knowledge sharing," in *Communities and Technologies*, M. Huysmann, E. Wenger, and V. Wulf, Eds. Kluwer Academic Publishers, 2003.

[9] G. Probst, S. Raub, and K. Romhardt, *Wissen managen*, 4th ed. Gabler Verlag, 1997.

[10] T. H. Davenport and L. Prusak, *Working Knowledge*, 2000th ed. Harvard Business School Press, 1998.

[11] B. A. Nardi, *Studying context: a comparison of activity theory, situated action models, and distributed cognition*. Cambridge, MA, USA: Massachusetts Institute of Technology, 1995, pp. 69–102. [Online]. Available: http://dl.acm.org/citation.cfm?id=223826.223830

[12] A. Kohlhase and M. Kohlhase, "Compensating the computational bias of spreadsheets with MKM techniques," in *Intelligent Computer Mathematics*. Springer, 2009, pp. 357–372.

[13] R. Mittermeir and M. Clermont, "Finding high-level structures in spreadsheet programs," in *Reverse Engineering, 2002. Proceedings. Ninth Working Conference on*. IEEE, 2002, pp. 221–232.

[14] B. A. Nardi and J. R. Miller, "Twinkling lights and nested loops: Distributed problem solving and spreadsheet development," *International Journal of Man-Machine Studies*, vol. 34, no. 2, pp. 161–184, 1991.

[15] D. G. Hendry and T. R. G. Green, "Cogmap: a visual description language for spreadsheets," *Journal of Visual Languages & Computing*, vol. 4, no. 1, pp. 35–54, 1993.

[16] T. R. Green, D. J. Gilmore, B. Blumenthal, S. Davies, and R. Winder, "Towards a cognitive browser for oops," *International Journal of Human-Computer Interaction*, vol. 4, no. 1, pp. 1–34, 1992.

[17] K. Wolstencroft, S. Owen, M. Horridge, O. K. O, W. Mueller, J. Snoep, F. du Preez, and C. Goble, "Rightfield: Embedding ontology annotation in spreadsheets," *Bioinformatics*, vol. 24, no. 14, pp. 2021–2022, 2011.

[18] A. Kohlhase, "Human-spreadsheet interaction," in *Human-Computer Interaction–INTERACT 2013*. Springer, 2013, pp. 571–578.

[19] A. Guseva, "Towards understanding context dimensions of spreadsheet knowledge," 2013.

[20] N. Dahlbäck, A. Jönsson, and L. Ahrenberg, "Wizard of oz studies: Why and how," *Knowledge-based systems*, vol. 6, no. 4, pp. 258–266, 1993.

[21] D. Spencer, *Card sorting: Designing usable categories*. Rosenfeld Media, 2009.

[22] M. G. Capra, "Factor analysis of card sort data: an alternative to hierarchical cluster analysis," in *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 49, no. 5. SAGE Publications, 2005, pp. 691–695.

[23] J. M. Corbin and A. Strauss, "Grounded theory research: Procedures, canons, and evaluative criteria," *Qualitative sociology*, vol. 13, no. 1, pp. 3–21, 1990.

[24] A. Salmoni. (2012) Open card sort analysis 101. [Online]. Available: http://www.uxbooth.com/articles/open-card-sort-analysis-101/