

SpreadComp platform: a new paradigm for distributed spreadsheet collaboration and composition

Simone Mangiante, Massimo Maresca
Computer Platform Research Center (CIPI)
University of Genoa, Italy

s.mangiante@cipi.unige.it, massimo.maresca@unige.it

Leonardo Roncarolo
M3S S.r.l.
Genoa, Italy

leonardo.roncarolo@gmail.com

Abstract—A huge amount of data is everyday produced, even in many critical business sectors, in a distributed, collaborative way with the support of spreadsheet applications. The manual fusion of such data often yields errors: copy-paste operations, loss of alignment and coherency due to multiple spreadsheet copies or broken cross-spreadsheet links. These problems are partially addressed by two of the principal software platforms for collaboration: Microsoft SharePoint and Google Docs. After describing features and tools of the two platforms specifically designed for spreadsheet collaboration, pointing out differences and limitations, we propose SpreadComp, a spreadsheet composition platform based on three new collaboration patterns which span users and data interaction from the simplest flat scenario to more structured and hierarchical ones. We discuss the positioning of our collaboration patterns with respect to the analyzed existing approaches and we present the SpreadComp architecture and its prototype implementation.

Keywords: *spreadsheet, collaboration, distributed, platform*

I. INTRODUCTION

Companies and organizations in many critical business areas base many of their most critical processes on spreadsheet applications [1][2][3][4]. Some of these risks are due to bad practices or lack of control in the development phase of spreadsheet applications by the End Users [5], others are hidden in the run-time processing of data contained in distributed spreadsheets [6]. Furthermore, in large and small organizations spreadsheet-based data processing often takes place in a distributed, collaborative way, where many actors produce or consume spreadsheet-based data and contribute to the setup of a real spreadsheet-based “distributed workflow”. The propagation and the update of data in these workflows are commonly performed either using the tools provided by spreadsheet applications (like cross-spreadsheet links) or using manual techniques, like copy and paste operations. Access to spreadsheet data is often not controlled, allowing different methods like concurrent access by different actors to a shared location or circulation of spreadsheets by email.

The combinations of these practices may lead to errors and/or inefficiencies, the most common of which are:

- inconsistent cross-spreadsheet links due to uncontrolled delete/rename/replace operations of spreadsheet files potentially leads to data loss;
- proliferation of many copies of data due to spreadsheet circulation by email or other means leads

to data replication, lack of control on spreadsheet versions and potential usage of obsolete data;

- uncontrolled concurrent access to a shared spreadsheet for manual data update leads both to high latency time for update (only one actor at time is allowed to modify data) and to uncontrolled changes in the spreadsheet data;
- manual copy/paste operations for propagating data from one spreadsheet to another leads to mistyping errors and violations of cell formats.

In order to provide a more controlled approach to distributed spreadsheet processing, we analyze and compare three IT environments which include features about spreadsheets, with focus on Microsoft Excel application:

- Microsoft SharePoint 2010 + Microsoft Office Excel client, with Excel Services component
- Microsoft Office365
- Google Docs

All the solutions are aimed at the corporate world and try to extend the capability of desktop system by adding group sharing and collaboration features.

The Microsoft Excel focus is motivated by a market research by Forrester [7] claiming that 81% of the surveyed organizations ran Office 2007, including Excel 2007, 78% supported by SharePoint.

We explore problems and patterns not covered by them and propose our solution, the SpreadComp platform. The paper is organized as follows: in Section II we describe investigated platforms, in Section III we focus on features of these platform designed to manage Excel documents, in Section IV we propose three new spreadsheet data collaboration patterns, in Section V we present the SpreadComp platform and in Section VI we provide some concluding remarks.

II. OVERVIEW OF THE INVESTIGATED PLATFORMS

A. Microsoft SharePoint 2010

Microsoft SharePoint 2010 [8] is a web application platform for business collaboration aimed at the enterprise market. It has a service oriented architecture (SOA), with the possibility to be scaled down to operate from a standalone

server or scaled up across several machines. To exploit SharePoint full capabilities clients (PCs) must be equipped with Microsoft Office suite. As an added bonus, a subset of SharePoint services is also available via web browsers, requiring no Microsoft product installed on the clients

SharePoint covers six capability areas [9]:

- Sites: a single infrastructure for every business website, from a project team site to a internet portal used by the organization clients; this area covers the Web Content Management System
- Communities: collaboration tools to help people within organizations to work together; this area includes tools like chat, wiki, blog, calendar, notifications, email integration and social connections
- Search: powerful search engine based on keywords, refinement content analysis and social clues helping people to find information in or outside organizations
- Content: management of documents and items needed in collaborations, including shared repositories, access control policies, rich media management; this area is about Enterprise Content Management and document management
- Insights: aggregation and elaboration of information from within organizations and the internet; this area is about Business Intelligence
- Composites: tools and components to make custom business solutions with no-code integration of data, documents and process; this area is about Workflows, Data Connectivity, Mashups

About security and accounting, users of Microsoft SharePoint 2010 must be authenticated and authorized to use its services. Registration takes place directly on the server or through an Active Directory or LDAP system. According to a survey done by AIIM in 2011 [10], SharePoint is adopted by 70% of largest organizations and only 20% of the total number of interviewed organizations has no interest in it. We focus on Microsoft SharePoint 2010 Enterprise Edition since it includes a component specifically addressed to Excel document management and business intelligence named Excel Services [11].

It provides:

- web publishing of an Excel document authored by a Microsoft Excel client
- recalculation of an Excel document on the server
- distribution of an Excel document through a web browser (without the need of Microsoft Office Excel client) or programmatically through a web service interface

Its architecture [12] is depicted in Figure 1. Its three main components are:

- Excel Calculation Server (ECS): a web service resident on a SharePoint server providing the functionality to open, close and perform operations on Excel spreadsheets

- Excel Web Access (EWA): it is a SharePoint web part, a Microsoft term for web widgets which can be embedded in other SharePoint web pages or dashboards; it translates an Excel document into a web page
- Excel Web Services API: it exposes a web service interface to Excel Services

B. Microsoft Office365

Office365 [13] is the Office solution provided by Microsoft as an answer to the developing market of cloud based solutions.

It is the SaaS deployment of Microsoft Office suite, commercially available through monthly or periodic subscriptions. It includes the following tools:

- Microsoft Exchange Online for email and scheduling
- Microsoft SharePoint Online for sharing files and creating team sites; it differs from SharePoint 2010 since lacks some advanced features for external data connection and custom third parties development
- Microsoft Lync Online for instant messaging and online meeting
- Microsoft Office Web Apps for lightweight editing of Microsoft Office documents inside a web browser; it differs from Microsoft Office desktop suites since lacks some advanced features only available in the desktop version

Microsoft Office365 runs entirely in the cloud. There is therefore no need for a standalone server to deploy.

Microsoft Office365 users must purchase a subscription to the cloud service and can share documents with other users external to the platform who have a Microsoft Hotmail or Live account.

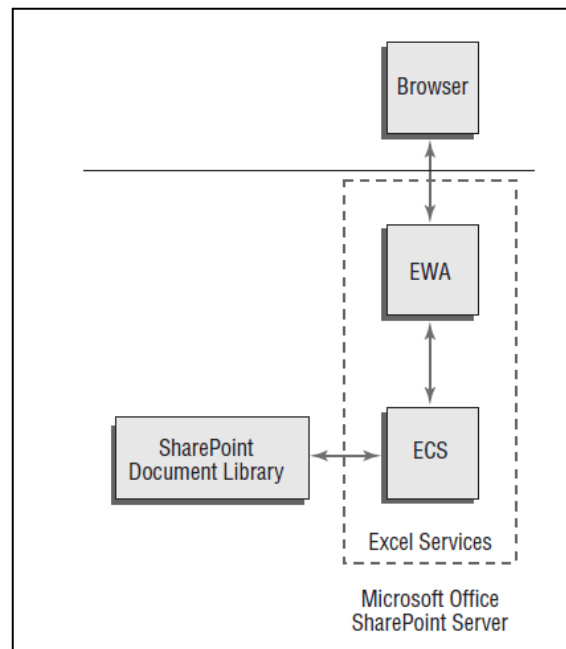


Figure 1. Excel Services architecture

C. Google Docs

Google Docs [14] is Google's SaaS office suite, delivered both free or through periodic payment subscriptions for the offering Google Apps. It allows users to create, edit documents online inside a web browser while collaborating in real-time with other users and store them in a cloud storage space managed by Google Drive [15] file storage and sync service. It is fully integrated with other Google's tools like Gmail, Google+ and Google Maps in order to extend collaboration to social media and other sources. It includes Google Spreadsheets to manage spreadsheet documents. Spreadsheets are stored in the Google Spreadsheets format but users can upload Excel formatted documents and Google platform can convert them (or maintained in their original format that cannot be edited). It also exposes a web service interfaces (Google Spreadsheets API and Google Documents List API, version 3.0 [16]) to access and manage documents programmatically.

Users of Google Docs must have a Google account and can invite anyone else: if the invited user doesn't have a Google account, he or she is forced to register one.

III. HOW THE INVESTIGATED PLATFORMS MANAGE EXCEL DOCUMENTS

The three environments provide features to manage the entire lifecycle of documents, with some differences and peculiarities we will examine for end users who needs creating and sharing Excel spreadsheets.

A. Authoring Excel documents

1) Microsoft SharePoint 2010 standalone + Microsoft Office Excel Client

Usually a user authors an Excel spreadsheet using the Microsoft Office Excel client, locally on his or her machine. After the creation of the spreadsheet, the author can proceed to the sharing phase.

2) Microsoft Office365 + Microsoft SharePoint Online

It offers the possibility to author a new spreadsheet directly through its web browser interface, or to use Microsoft Office Excel client.

3) Google Docs

It provides a web browser interface to author spreadsheets and immediately store it in the cloud in the Google Spreadsheets format. Microsoft Office Excel client can be used but it needs the Google Cloud Connect add-in [17] to connect and synchronize documents from the local machine to the cloud.

B. Sharing Excel documents

1) Microsoft SharePoint 2010 standalone

Microsoft SharePoint 2010 has an Enterprise Content Management functional area designed to host and share documents (included Excel spreadsheets). A user can create document libraries and internal or external sites to have a common repository for sharing documents with a group of users. Once the document is stored in the repository, the creator can organize the structure of the repository, decide access and control policies, bind the file to workflows and rely on other tools automatically provided by the platform like alerts,

logging, auditing and versioning. The creator can organize users to share the document in groups and assign them permissions to read, modify and download documents.

Microsoft SharePoint 2010 provides a powerful versioning engine which can manage major and minor version numbers (to distinguish content modifications from small grammatical changes), let users decide how many versions to maintain on the server, expose check-out and check-in phases to allow users to concurrently work on the same document, like in CVS [18] and SVN [19] systems.

A workflow [20] engine, based on Workflow Foundation [21], is also available. It manages three types of workflows:

- Out-of-the-box: a predefined common workflow ready to run without coding; SharePoint includes several out-of-the-box workflows about document approval, feedback and signatures collection, issue tracking
- SharePoint Designer: a workflow defined through the SharePoint Designer tools, which allows customization in a simple graphical way without particular programming skills
- Visual Studio 2010: a fully custom workflow coded in .NET framework in Visual Studio

2) Microsoft Office365 + Microsoft SharePoint Online

In this environment users can do the same operations as in the SharePoint 2010 standalone environment, since SharePoint Online doesn't differ from SharePoint 2010 with respect to the features described above. There is no need to upload anything because spreadsheets created with Excel Web App are automatically saved in the document library chosen by the author.

3) Google Docs

When an author creates the spreadsheet through the Google Docs web interface the spreadsheet is stored in the Google Spreadsheets format, while when the author upload an existing local Excel spreadsheet he or she can decide to convert it to the Google Spreadsheets format; conversion is mandatory to enable the real-time collaboration on the document, otherwise users can only view and must download to their PCs to edit the files. The author can choose the visibility of the uploaded spreadsheet and assign permissions to people (both singular user and mailing lists) whom sharing the document with. Google Docs integrates a light versioning system called Revision History which keeps track of every modification done to the document by enabled users. The author can create folders in the cloud storage space, but these operations have only a user space scope, there is no possibility to organize the storage space for every user participating in the document sharing, as in common remote repositories. In Google Docs the author can also set notifications to find out when collaborators have modified the spreadsheet, defining a set of rules which drive the notification engine.

C. Web publishing of Excel documents

1) Microsoft SharePoint 2010 standalone

The Excel Services component allows users to publish an Excel spreadsheet in the web. If managed by this tool, an Excel

file stored in a SharePoint document library can be accessed via web browser by users who have permission to view it, since the EWA module transforms it in a HTML web page almost identical to the Microsoft Excel desktop client interface. The owner of a spreadsheet document can be interested in sharing only a part of the document. It typically occurs when a complex spreadsheet must be shared with other users who, need accessing only the output data. This sharing mode is partially managed by the Microsoft SharePoint Excel Services component. The publisher can decide to publish only a range inside a spreadsheet, but Excel Services needs to store the entire file and annotates it with the definition of the range to publish. When a user has no permission to access an Excel spreadsheet through the EWA, he cannot modify the file, nor view the formulas contained in the sheet. The user can navigate, sort, filter the data and refresh calculations thanks to the ECS module. If the publisher defines special blank cells inside the spreadsheet, called parameters, viewers are allowed to insert data in those cells and recalculate the spreadsheet. If the user has the permission to download the file, he or she can download a copy locally on his personal machine and make changes. The master copy remains the one stored on the server, and other users who access the EWA to that file can't see the changes. If the user who locally modified the file has the permission to upload a new version of that file, he can update the master copy on the server and thus commit changes.

2) *Microsoft Office365 + Microsoft SharePoint Online*

This environment behaves as the SharePoint 2010 standalone one, because the Excel Services component is included also in SharePoint Online.

3) *Google Docs*

Google Docs provides a web publishing feature: a user can publish a spreadsheet on the web and get its link to send to other users. When a spreadsheet file is published on the web in Google Docs, the publisher can't define access control policies: the spreadsheet is thus public, anyone who receives the link can view or download the file. The publisher can generate different links which points to different formats (HTML web page, HTML snippet, CSV file, TXT file, PDF file, ATOM or RSS feed, XLS file, ODS file) of the published spreadsheet and share them on the web or by email. In Google Docs a user can only share an entire file, therefore the minimum unit of sharing for a spreadsheet document is the entire spreadsheet itself. The only exception is when a spreadsheet is published on the web: in this case the owner of the document can choose to publish a single named range on a sheet, thus the minimum unit of sharing becomes a cell range.

D. *Collaborating on Excel documents*

An author of a spreadsheet usually share a document not only to make it accessible by other users in a simpler way, but also to start a collaboration with other users, in order to build a richer and more complete work product.

With regard to concurrency, it can be:

- Real-time, when users modify a document simultaneously
- Non-real-time, when users must lock the document to apply changes to it

1) *Microsoft SharePoint 2010 standalone*

It offers only the non-real-time “one file – many users” collaboration. Only one user at a time can modify a document. Users must check the spreadsheet out of the document library on the server in order to make changes to it. The file is then saved on the user's local machine, modifications can only be done through a Microsoft Office Excel client. Then the collaborator who has checked the file out must check it in to commit modifications, as if he or she uploads the file for the first time. SharePoint 2010 provides two types of check-in and check-out mechanism:

- Short-term (default behavior): a user can modify a file without explicitly do a check-out, but for a limited period of time (10 minutes in Excel 2003, 60 minutes in Excel 2007 and 2010), after that Excel client requests additional time if it is open or the file is check-in automatically by SharePoint
- Long-term (can be enabled on documents and libraries): a user must explicitly check-out the document (through Excel client or SharePoint web application) and, after modifications, even if Excel client is closed, he or she must manually check the document in

It provides also a feature called Document Workspace, a specific type of site where a single document is the center of attention. Using this specialized document library, collaborators can work on a single file viewing in a compact manner, also inside Microsoft Excel client, previous versions of the document, the other users involved, tasks relating to the document, announcements and notifications lists.

2) *Microsoft Office365 + Microsoft SharePoint Online*

This environment has the features of the standalone product, thanks to SharePoint Online, but adds the real-time collaboration, available for spreadsheet only using the Excel web application from the web browser: users can edit an Excel document simultaneously through the Excel web application inside a web browser and the Office365 cloud platform synchronizes in real time every modifications, updating the document view of every user.

3) *Google Docs*

Google Docs is designed to implement the real-time collaboration, providing the same feature as Microsoft Office365, without the Excel desktop client's look and feel it shows. Google Docs, relying on other web applications in the Google cloud environment, provides other tools to support collaborations: email notifications, chat, blogs, calendar and offline synchronization of the document. The non-real-time collaboration in Google Docs is implemented without a check-out/check-in mechanism: users can download a copy of the spreadsheet to the Microsoft Excel client with the Google Cloud Connect add-in, edit it and synchronize changes with the cloud.

E. *Summary comparison*

TABLE I. summarizes the spreadsheet related features implemented by the three environments, including a comparison with the SpreadComp features described in the following sections.

IV. NEW SPREADSHEET DATA COLLABORATION PATTERNS

According to existing studies about spreadsheets utilization and adoption [22][23], we noticed that spreadsheet sharing, collaboration and composition can be more intuitive and simpler (without compromising security) than their implementations on the three analyzed environments. Spreadsheet authors should not work with other tools than spreadsheet applications (in our prototype Microsoft Excel but the approach is general and there is no restriction on the spreadsheet application used providing it supports extensions), they should share and exchange tables of data remaining always inside the spreadsheet application, without configuring or invoking other systems manually. Furthermore a new paradigm encourages an approach derived from software development practices: definition of I/O interfaces, modularity, hiding, definition of collaboration in a more formal way, etc. From these needs we developed our SpreadComp platform.

We found three main collaboration patterns, suggested by three data sharing scenarios gradually more structured and more hierarchical.

A. Flat

This is the simplest scenario, in which a user needs to share data from one of his spreadsheets with other users through an internet network.

From the user point of view, sharing takes place simply by selecting the area(s) to be shared (becoming a data exporter) and making it available to other users (who become data importers). Importer users get a notification of the sharing event and can (from the spreadsheet application) insert the data area in their file. The file is automatically linked to the source and will be updated every time the source changes. Only data get exported. Receivers cannot see the formulae that produced the data. The other two collaboration patterns are built on top of this basic mode. This is the collaboration pattern at the core of our first distributed spreadsheet composition platform, named DISCOM [24], which now has been extended and has become SpreadComp.

B. Reverse Peer To Peer (RP2P)

In this scenario a user “requests” other users to provide data according to a predefined pattern (template). A template is a fixed styled and formatted spreadsheet area (with formulae) a user provides to request spreadsheet data. The receiving user(s) fill the received template with data and send back the evaluated results. Inside a typical spreadsheet application a template is therefore a cell range which can contain:

- Text cells, usually associated to heading columns/rows or cells fixed by the requester user
- Formulae which must refer only cells belonging to that range
- Blank cells where users put their data

We call this interaction, which is more structured than the previous one, Reverse Peer to Peer (RP2P). It consists of these steps, depicted in Figure 2:

1. A user U1 defines a template through which to request data to other users, called target users;

2. the target users receive a notification from the platform about the template created by U1 and have to accept and install the template to start data sharing

These steps rely on the previous collaboration mode, since they can be transformed as follows:

1. U1 becomes an exporter of the template and an importer of the template installed and filled by target users
2. Target users become importers of the template created by U1 and exporters of the template filled with their data

C. Composition

In this scenario a user wants to establish a complex collaboration involving several users and/or templates, defining a priori who produces which spreadsheet data and who consumes it. This collaboration pattern is similar to a data workflow, since the life cycle of data is well defined: it is more structured than the RP2P interaction and can resemble the hierarchical environment usually characterizing enterprises. The SpreadComp platform provides the required tools to design collaborations, which we call Composition, preferably in a graphical way, and translate its abstract outcome in many interactions of the previous type. This collaboration pattern is helpful in organization domains, where a coordinator/designer knows how to build specific internal spreadsheet data and can design a fixed Composition to accomplish information gathering for complex tasks.

V. ARCHITECTURE OF THE SPREADCOMP PLATFORM

The proposed spreadsheet data collaboration patterns are supported by our platform, named SpreadComp, based on a client-server architecture: one or more clients interact with the server platform across a network, which can be an Intranet, a VPN or the Internet.

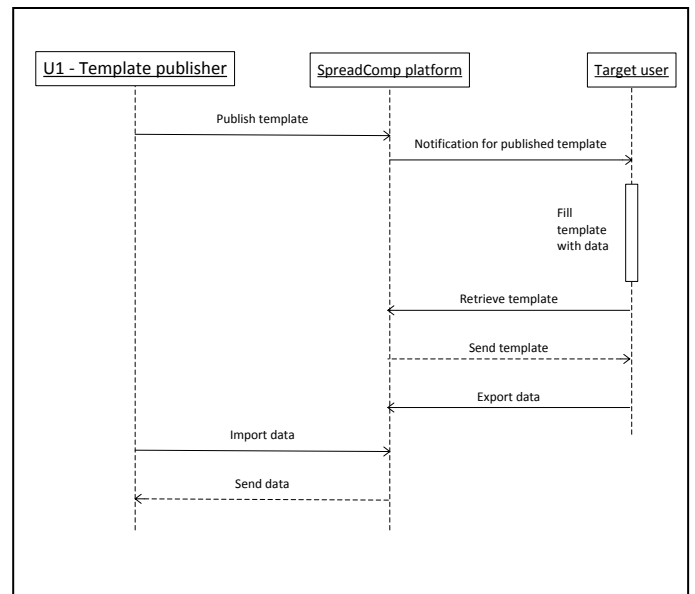


Figure 2. Sequence diagram of a template publication

The client part is an Add-in module integrated in the client's local Microsoft Excel desktop application. The module interacts with the composition platform by means of Web Services [25] and allows the user to configure itself and make exports and imports of spreadsheet areas through a graphical interface. The SpreadComp platform exposes interfaces that support the execution of spreadsheet data exports and imports, manage user accounts and synchronize data across connected spreadsheets. The platform architecture is depicted in Figure 3: the two main components are the Add-in Client and the SpreadComp Server. We adopted an architecture based on a central composition server in order to provide a central repository that performs automatic synchronization even when one or more spreadsheets in the workflow are offline (by delegating to the server the task to produce the results). The system comprises two additional entities: the graphic composition designer and the monitoring tool.

A. Client Add-In

It is written in C# using Visual Studio Tool for Office [26]. It provides a graphical interface that allows users to control the import/export operations; the interface maintains the same look and feel as the other Microsoft Excel functions, with respect to different versions of UI layout (command bar or ribbon interface). The main design objective was to integrate it as seamlessly as possible in the Excel application, letting users concentrate on normal Excel commands without worrying about other tools. The Add-in interacts with the SpreadComp Platform through Web Services technology, as Web Services are widely developed and supported in enterprise environments. In particular we adopt standard SOAP messaging. The cell range to be exported is converted to a custom XML document and wrapped in a SOAP request which is sent to the SpreadComp Platform. The SpreadComp Platform module stores the XML translation of the exported cells in the database.

The Add-in is designed to run in single user mode, requesting login credentials at startup. Status information is stored on the client machine:

- exports and imports information are kept in the spreadsheet, inside file custom properties;
- user identification, SpreadComp server address and other Add-in configuration data are kept in user's local space on the filesystem.

B. SpreadComp Server

The server side of the platform is written in Java, using Spring framework [27] as the engine for managing Web Services and reacting to HTTP and SOAP requests. The main interface exposed to the Client Add-In is a Web Service which provides a set of functions that allow clients to authenticate themselves and perform CRUD (Create, Read, Update, Delete) operations on exported and imported data. It serves client SOAP requests by invoking Spring beans interacting with a MySQL database [28] in a transactional environment.

The server includes a spreadsheet engine: when a spreadsheet file declares automatic exports depending on automatic imports it must be uploaded to the server to be reevaluated when imported data change even if it is not open in

the Excel desktop client where it was created. The engine is a Spring bean which performs this sequence of operations:

- 1) detection of the intermediate uploaded file to be re-evaluated;
- 2) insertion of updated values in its automatic imports;
- 3) recalculation of cell formulas;
- 4) update of automatic exports (those modified by the recalculation).

Every step is performed using Apache POI library [29], a set of objects able to open and manipulate Microsoft Excel file directly within Java code.

For administration purposes a web application is provided by the platform. Through this component an administrator can manage user accounts and control the platform working correctly.

The platform can be deployed in two different ways:

- as an enterprise service, creating a "ad-hoc" server inside an enterprise domain to give the enterprise full control over the system and limit spreadsheet data circulation within the enterprise domain; we call this deployment mode Corporate
- as a cloud service hosted on a IaaS (Infrastructure-as-a-Service) provider like Amazon Machine Image – AMI on the Amazon EC2 Infrastructure [30], to enable an easier and quicker setup for spreadsheet compositions which don't need strong control and data flowing-over-the-internet limitations; we call this deployment mode Open/Cloud

C. Logical units

Figure 4 depicts the logical units of the SpreadComp platform with respect of the following four layers:

- The communication layer provides the link between clients and server; it is the internet network and uses HTML and SOAP protocols to expose the platform web services

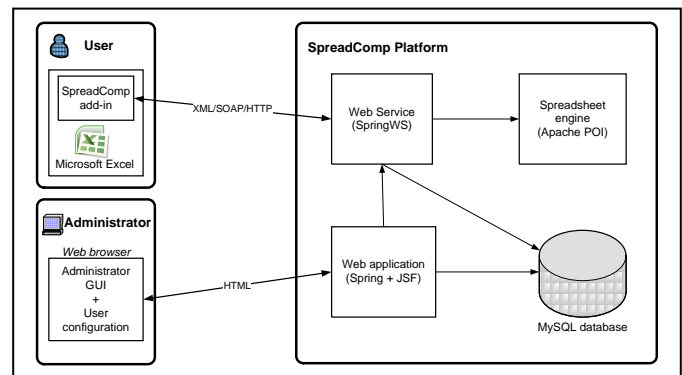


Figure 3. SpreadComp architecture

- The basic layer provides the functionalities required to implement the collaboration patterns described in Section IV
- The collaboration layer provides access to the basic services and methods to involve other users in the collaboration
- The application layer offers service access and configuration facilities to the platform through the Add-in client and the other external tools

Authentication and access control provides a secure access to the SpreadComp server, it is the first point of access for every operation made on the platform either via web service methods or web application. It is implemented as a validation of a couple of strings username/hashed password over a HTTPS channel.

Basic Export/Import realizes the Flat collaboration pattern.

The *Templates* component realizes the Templates collaboration pattern. The Add-in exposes two operations:

- Publication of a cell range as a Template: in this case not only the data and the number formats are sent to the platform, but also formulas and styles contained in the range, with the strict constraint that nothing can reference any cell or sheet outside the range
- Installation of an existing template on the platform by a target user: this operation imports the template on a Microsoft Excel sheet, with formulas and styles as defined by its publisher, and instantly creates an automatic export of only the data and number formats contained in the same range for the template publisher

In this way the template collaboration pattern is implemented relying on the flat collaboration pattern and it is ensured that the data requested by the template publisher is always updated whenever a target user fills the template with new data.

The *Compositions* module is integrated with the *Composition Designer* external tool. It is a standalone graphical Windows program by which users can design a spreadsheet Composition collaboration and deploy it inside the platform. Every platform user is represented by a graphical block and can be linked to another user by a link which represents the RP2P interaction between them. The arrow is annotated with the name of the template through which data will be shared (the arrowhead points to the template publisher). When the Composition designer, who must be a platform user but doesn't need to participate in the Composition itself, deploys the Composition to the platform, the SpreadComp server triggers a RP2P interaction for every arrow in the Composition using the Templates logic module.

In the example depicted in Figure 5, taken from the real Composition Designer tool, during the deployment phase the user B receives a notification informing him there is the template T to install and the user A becomes the importer of the data filled in T by B.

Every Composition id stored on the SpreadComp server in a XML format and can be retrieved from the server in order to be eventually updated or modified by its creator. The transformation from the graphical representation to the RP2P interactions to instantiate is made by the SpreadComp server, so the Composition Designer tool doesn't need to synchronize anything locally. The only mandatory requirement is a network connection to the SpreadComp server when the Composition is deployed.

The *Notification* component is concerned with informing the platform users whenever something needs being notified to them. For example, if a user is a target for a new data export or a new template. In our prototype the notifications are simply email sent from the SpreadComp server to the user, who needs to give his or her email address to the platform during the registration phase in the Open/Cloud version or the system setup in the Corporate version.

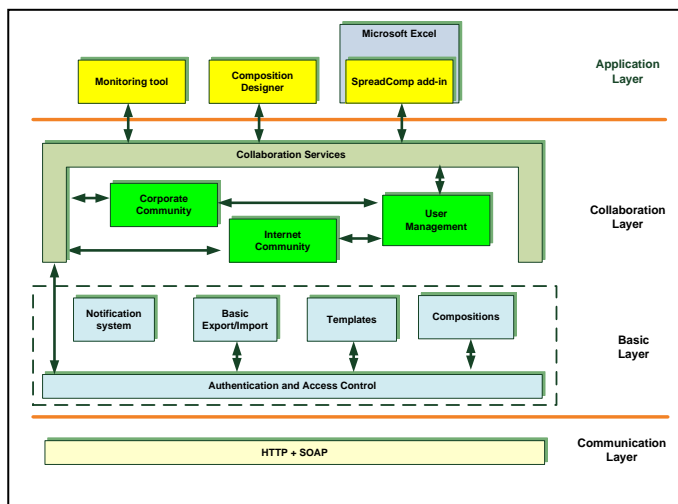


Figure 4. SpreadComp logical units

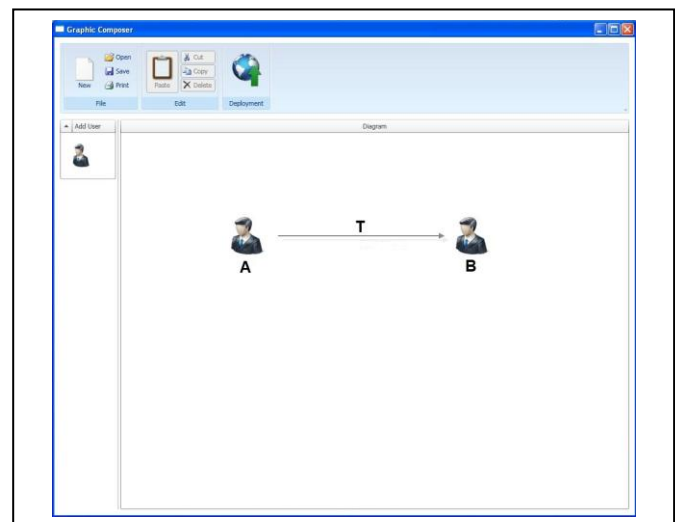


Figure 5. Composition Designer tool

User Management is a module inside the SpreadComp server that manages platform users and their relationships with other users and spaces. It can use two different components, with respect of Corporate or Open/Cloud deployment of the system:

- *Corporate Community*, which creates users inside an organization domain under the supervision of a system administrator
- *Internet Community*, which create users over the internet through a web registration form and a viral invitation mechanism transported by email notifications

Monitoring tool is a Windows standalone program that shows the status of all the data passing through the platform. A platform user can view all his or her interactions, in a tabular form filtered by several features, and check their status. A user with administrator privileges can also view every interaction stored in the SpreadComp server along with some log files.

VI. CONCLUSION

We have analyzed existing platforms managing sharing of Excel documents and collaboration on them.

We have proposed three new spreadsheet data collaboration patterns, emerged from common organizations' scenarios, we didn't find in the investigated platforms.

We have presented a new platform which implements these patterns, called SpreadComp.

We have compared the features implemented in SpreadComp with the analyzed platforms, stressing the ease of use and the simplicity, as summarized in TABLE I.

A prototype of our platform was developed to demonstrate the concept. We are planning to run trials and experimentations in order to evaluate the platform features and the usability of our approach.

The objective is to refine the platform specifications according to suggestions from test users, to add missing feature already planned (like a versioning system), to update the software and make a stable deployable product.

[1] The TABB Group, "Spreadsheets and Capital Markets: The State of the Enterprise", <http://download.microsoft.com/download/4/6/1/461062F0-2D3E-4620-890E-826CE55D27B9/SpreadsheetsandCapitalMarkets.pdf>, 2009

[2] A. Kelsden, "Narrowing in on spreadsheet", Pharmtech.com, <http://pharmtech.findpharma.com/pharmtech/Regulatory+Report/Narrowing-in-on-spreadsheets/ArticleStandard/Article/detail/319885>, 2009

[3] EuSprig, "Horror Stories", <http://www.eusprig.org/stories.html>, 2010

[4] G.J. Croll, "Spreadsheets and the Financial Collapse", EuSprig Conference 2009

[5] S. Saadat, "Managing Linked Spreadsheets in a Compliant Environment", EuSprig Conference 2007

[6] D. D. Hoang et al, "An analysis of spreadsheet-based services mashup, Proceedings of the Twenty-First Australasian Conference on Database Technologies" - Volume 104 (Brisbane, Australia, January 01 - 01, 2010). H. T. Shen and A. Bouguettaya, Eds. Conferences in Research and Practice in Information Technology Series. Australian Computer Society, Darlinghurst, Australia, 141-150, 2010

[7] <http://www.zdnet.com/blog/microsoft/forrester-google-still-a-distant-office-competitor/6168>

[8] C. Geier et al., "SharePoint 2010 Six-in-One", Wrox Programmer to Programmer, 2011

[9] Microsoft, "SharePoint 2010 evaluation guide", 2010

[10] AIIM, "State of the ECM Industry 2011", 2011

[11] S. Prish, "Professional Excel Services", Wrox, 2007

[12] L. Asnash, E. Megiddo, C. Thomas, "Beginning Excel Services", Wrox, 2007

[13] K. Murray, "Office365. Connect, collaborate virtually anywhere, anytime", Microsoft Press, 2011

[14] http://en.wikipedia.org/wiki/Google_Docs

[15] <https://drive.google.com/>

[16] <https://developers.google.com/google-apps/documents-list/>

[17] <https://tools.google.com/dlpage/cloudconnect/>

[18] http://en.wikipedia.org/wiki/Concurrent_Versions_System

[19] http://en.wikipedia.org/wiki/Apache_Subversion

[20] D. Hollingsworth, "Workflow Management Coalition. The Workflow Reference Model", Workflow Management Coalition, 1995

[21] http://en.wikipedia.org/wiki/Windows_Workflow_Foundation

[22] K. R. Baker, L. Foster-Johnson, B. Lawson, S. G. Powell, "A Survey of MBA Spreadsheet Users", Tuck School of Business, 2006

[23] S.E. Kruck, "Towards a theory of spreadsheet accuracy: an empirical study", unpublished, 1998

[24] P. Baglietto et al., "Always-on distributed spreadsheet Mashups", Proceedings Mashups 09/10, 2010

[25] G. Alonso et al., "Web Services", Springer Verlag, 2003

[26] [http://msdn.microsoft.com/it-it/vsto/default\(en-us\).aspx](http://msdn.microsoft.com/it-it/vsto/default(en-us).aspx)

[27] <http://www.springsource.org/>

[28] <http://www.mysql.it/>

[29] <http://poi.apache.org/>

[30] <http://aws.amazon.com/ec2/>

TABLE I. SUMMARY COMPARISON OF ANALYZED FEATURES IN THE THREE ENVIRONMENTS

Feature	Environment			
	<i>Microsoft SharePoint 2010 standalone + Microsoft Office Excel Client</i>	<i>Microsoft Office365 + Microsoft SharePoint Online</i>	<i>Google Docs</i>	<i>SpreadComp platform</i>
Microsoft Excel client needed for authoring	Yes	No	No	Yes
Accounting	Users must be registered in SharePoint 2010 or in a Active Directory/LDAP environment	Users must be registered in SharePoint Online or invited if they have a Hotmail or Windows Live account	Users must have a Google account.	Users must only have a valid email address
Authoring via web browser	No	Yes	Yes	No
Structured and customizable repositories	Yes	Yes	No, flat cloud space with user defined folders	No
Document access control policies	Permissions set for users and groups	Permissions set for users and groups	Permissions set for users and mailing lists	Spreadsheets stored only to manage chains, not available to user
Versioning	Yes, with major and minor number and check-in/check-out mechanism	Yes, with major and minor number and check-in/check-out mechanism	Yes, it keeps track of recent document versions	No. Planned feature for tracking exported spreadsheet modifications
Workflows	Yes	Yes	No	Basic, implemented as Compositions
Notifications	Yes	Yes	Yes	Yes, only by email
Control on who can view a publication	Yes, as in document sharing situation	Yes, as in document sharing situation	No, everyone having its link can view it	Yes
Sharing part of a document	Yes, but the entire document must be uploaded	Yes, but the entire document must be uploaded	Yes, but the entire document must be uploaded	Yes
Users' interaction with a publication	Navigation, filter, sorting, parameters setting through a web browser with an interface identical to Microsoft Excel desktop client	Navigation, filter, sorting, parameters setting through a web browser with an interface identical to Microsoft Excel desktop client	View of HTML rendered tables through a web browser	Import, even multiple, in any spreadsheet with automatic or manual synchronization with the server
Possibility to download a copy of the publication	Yes, by users who have the right permission, only in Excel format	Yes, by users who have the right permission, only in Excel format	Yes, for everyone who has its link, in several formats	No, only imports. The document originating an exported spreadsheet area is not available to users
Minimum unit of sharing	Cell range	Cell range	Cell range	Cell range
Possibility to modify the spreadsheet originating the publication	For users with the right permission, only through a local copy obtained by checking the spreadsheet out	For users with the right permission, only through a local copy obtained by checking the spreadsheet out	No	No
Real-time collaboration	No	Yes	Yes	No, actually outside the scope of the project
Non-real-time collaboration	Yes, with check-in/check-out mechanism	Yes, with check-in/check-out mechanism	Yes, through Google Cloud Connect add-in for Microsoft Excel desktop client	Yes, through 3 new collaboration patterns
Other tools helping collaboration	Document workspace, wikis, blogs, instant messaging, email notifications, tasks	Document workspace, wikis, blogs, Microsoft Lync Online, email notifications, tasks,	Blogs, chat, email notifications, tasks, social connection to Google+	Graphical composition designer, monitoring tool