Integrating Workflow and Document Management with Domain Applications by Using MS SharePoint 2010 with REST Interfaces

Christopher Stolle, Kurt Sparber, Matthias Menger

Abstract
Public authorities need a way to integrate their executive domain applications with document and workflow management facilities. This article describes patterns of a solution based on MS SharePoint 2010 with REST interfaces, which is currently under development by Division IV of the Federal Environment Agency (Umweltbundesamt, UBA) in Germany.

1. Introduction
The need of Environment Agencies for data and information covers a wide range of topics. Moreover most of the topics are of overlapping nature, like energy, climate change, biodiversity or emissions to name just a few. The task of an information system is to cover the requirements of data and information on a particular topic for the target user group(s). Regarding the overlapping nature of all these topics all agencies have to deal with the problem of data redundancy and multiple systems storing, searching and retrieving this data and information. For user groups with aspects of scientific research this might end up in a chaotic assemblage of very heterogeneous systems and data.

In the following we are looking into the world of chemical safety which seems like a well-defined area at first but turns out to be very heterogeneous regarding the factual data needs and requirements for the specific legal enforcements (chemicals, biocides, pesticides, medicals).

Moreover, each legal jurisdiction has its own workflow-related data and information needs which have to be covered by the information system handling both factual and administrative workflow data. In 2009 the Federal Environment Agency kicked off a project to implement the concept of the redesigned Information System Chemical Safety (ICS), as the original version from 10 years ago is now somewhat dated.

The concept demands a split of the current system into separate components to better handle the respective data/information needs, i.e. factual and administrative data/information. This requires a component like a DMS and WMS handling workflows and administrative data objects (e.g. documents, letters, reports, studies, timelines, responsibilities etc.) with exact references to the factual data derived out of these data objects. Remembering the wide variety of topics in an environmental agency, each with its own workflow and data objects, the eagerness to establish a single system to handle the administrative data/information is quite obvious. The need to provide and share data and information across different departments more efficiently is a well-known target of any organisation. But regarding the already evolved and existing IT landscape this becomes a real challenge.

On the one hand Environmental Agencies usually operate a number of very heterogeneous domain applications due to their wide range of topics they have to deal with. On the other hand they are eager to establish a single tool for workflow and document management (W/DMS) to be shared by the various departments and their employees for improved efficiency of information provision and sharing.

1 innoQ Deutschland GmbH, Halskestr. 17, 40880 Ratingen, Germany {christopher.stolle|kurt.sparber}@innocq.com
2 Umweltbundesamt, Wörlitzer Platz 1, 06813 Dessau-Roßlau, Germany. matthias.menger@uba.de
• Workflows are based on activities to be executed with one of the domain applications and they have mostly documents or different data objects as input and/or output of these activities, so everything should be well integrated. Some examples:

• Starting a workflow activity in the WMS dialog should directly link to the dedicated section of the appropriate domain application and create links from there to the documents which reside in the DMS and are required as input.

• Resuming an activity should be started from the list of open issues in the WMS and directly link to the respective domain work place.

• Selected domain and W/DMS data should be replicated and synchronized between the involved systems, but there should be a single point of editing for each data item.

• Artefacts which are created during execution of an activity should automatically be uploaded to the DMS, and a link from the domain workspace to such artefacts should be created.

• The domain application workspace should link to the respective workflow activity so it is easy to adjust the state of completion.

• No redundancy of documents/data objects is required while serving different domain applications.

In most cases, such patterns of integration are very difficult to establish for three main reasons:

1. W/DMS has no convenient built-in interfaces to send and receive such requests

2. Domain applications are very heterogeneous so that each of them has to be handled individually when multiple interfaces are going to be implemented.

3. There is no established open standard (e.g. Wf-XML, CMIS, XDomea) which explicitly covers patterns as those described above.

The Federal Environment Agency currently works on an exemplary solution while implementing a new generation of a specific domain application i.e. Information System Chemical Safety (ICS, see also Menger et al. 2011) with an integration of a W/DMS which is based on MS SharePoint 2010 (Krause, Jörg et al. 2010). SharePoint 2010 widely fits these requirements with its internal document and workflow management system represented by a single web application.

For the appropriate use of this product, however, interoperability with other applications and the import of data and documents via interfaces is very important.

SharePoint 2010 offers several options for this requirement:

• Server-/Client Object Model

• Web services (SOAP)

• REST

Specific requirements from the project for the Federal Environment Agency are offered by the following points:

• Data transfer between SharePoint 2010 and the internal application used to manage chemical substances (ICS)

• Importing records and dossiers with the appropriate documents from different authorities with different exchange formats (XDOMEA, Caddy TOC)

The challenging aspects for the implementation are matching the requirements to the capabilities of SharePoint 2010 and develop extensions only when inevitable.

2. Key Concepts of MS SharePoint 2010 Integration

Due to the special requirements of the project for Germany’s Federal Environment Agency, SharePoint 2010 is used as document management and workflow management system. In addition, the SharePoint Solution has to be integrated with another web application – an information system for chemical substances (ICS). Out of the box SharePoint 2010 offers six key concepts for a potential approach:
• List and document libraries
• Lookup fields and managed metadata
• Workflows
• Web Parts
• Event receiver
• OData³/ REST⁴ Interface

2.1 List and document libraries
SharePoint lists and document libraries are used for data storage and can be compared with a table of a database. Like database tables, lists and libraries contain columns or fields with a specific type. Metadata for list entries is stored in these columns. Field types are almost equivalent to the types of columns or fields in a database or an object-oriented programming language. Apart from these field types there are two special types which are explained in the following section.

A document library is an extension of a SharePoint list where a document can be stored with its associated metadata. Multiple versions of a document can be managed at the same time.

2.2 Lookup fields and managed metadata
Lookup fields are used to create associations between two lists or rather their fields. Here again the comparison to the relational database model can be drawn. A lookup field of list A contains a reference to a field of list B. A specific entry of list A contains one or more values from the associated column in list B. Managed metadata describes an approach to central storage and management of a hierarchically structured list of terms that can be used as attributes for elements. Expressions have a unique ID and can have multiple different labels (synonyms). A collection of related terms is a term set. A term set can be created in two different scopes, a local scope in the context of a web page collection and in a global scope for all websites. In addition, one can define whether a term set is created open or closed. If the term set is defined as open, users can add new terms when updating a column which is assigned to this term set. Otherwise, this is only possible for users with specific management privileges of the central managed metadata storage. Permissions can be assigned to groups of term sets. In order to be able to use managed metadata in a list, there is a corresponding column or field type which refers to a specific term set.

2.3 Workflows
A workflow in the SharePoint context can be described as a predefined set of tasks whose execution results in an outcome. The tasks are stored in a special workflow list. Through the execution of a workflow, for example, field and list contents can be added, changed, moved or deleted and information from different users can be obtained at runtime. Workflows can trigger events, be held for a specified period of time and require user interaction for completing the user tasks. There are two different types of workflows, the configurable workflow and the code workflow. The configurable workflow can be created by privileged power users and administrators using the SharePoint Designer and offers support for simple processing.

⁴ Representational State Transfer (Fielding, R. T. (2000))
The SharePoint REST interface is not completely RESTful but follows most of the key constraints of the REST architectural style.
procedures. The processing is always performed sequentially. Due to the configuration approach changes at runtime are possible. Internally, a configurable workflow is built on an XML structure. A code workflow has to be developed in Visual Studio and is much more powerful than a configurable workflow. The implementation is quite complex and can only be performed by software developers. Changes at runtime are not possible. Code workflows are used for complicated processing procedures which require the implementation of a state machine.

2.4 Web Parts

Web Parts are used in the SharePoint context for reading or manipulating content on a SharePoint site by the end user. The changes can be made visible either to all or only for certain users. Furthermore, settings of a Web Part can be stored permanently for the user who made the changes. This makes it possible for end users to customize parts of a web page to their own needs.

Most Web Parts are used for the representation of SharePoint lists or libraries. For example, when a list is created in SharePoint, the representation can be displayed multiple times in a SharePoint website through the associated Web Part. A Web Part is automatically created for every new list or library. User permissions can be set in the scope of each Web Part.

2.5 Event Receiver

Event receivers offer the ability to respond to events which have been triggered inside a SharePoint application. An event, for example, is triggered when creating or modifying a list item. Event receivers consist of a code class and a configuration file in XML format. Inside the configuration file information is stored about which event in which scope a receiver should respond to. If an event which a receiver listens to is triggered, the code of its class is executed.

A complete list of the different events in SharePoint can be obtained from the following URL: http://msdn.microsoft.com/en-us/library/gg981880.aspx

2.6 OData/ REST Interface

In SharePoint 2010 for the first time a so-called REST Interface has been introduced to access lists and list items. This interface is based on WCF5 Data Service and the OData protocol. The WCF data service is not part of SharePoint development, so it usually has to be installed separately.

The Data service rather provides a general platform for accessing data sources using non-proprietary standards like:

- REST as the underlying architectural style
- HTTP used as transport or more accurately transfer protocol
- XML, JSON6 and RDF7 as the preferred data transport layer

OData is a Microsoft specification for a protocol for cross-platform integration of data. OData itself, as well as the WCF Data Service, is based on standards (HTTP, AtomPub8, JSON) and provides access to information from a variety of web applications and services. Microsoft released OData under the Open

---

5 “Windows Communication Framework” is part of Microsoft’s .NET Framework - http://msdn.microsoft.com/de-de/netframework/aa663324
6 “Java Script Object” - http://www.json.org
7 “Resource Description Framework” - http://www.w3.org/RDF

Copyright 2011 Shaker Verlag Aachen, ISBN: 978-3-8440-0451-9
Specification Promise (OSP). For this reason there are many frameworks and SDKs for a variety of pro-
gramming languages like C#, Visual Basic, Java, JavaScript, PHP, Ruby and Objective C.

3. Implementation of the Requirements

The use of the previously discussed concepts, supplemented by a few custom implementations, is suffi-
cient to fit the existing requirements. The explanations in the following sections describe the general pat-
terns implemented by SharePoint as a workflow and document management system and how to integrate
SharePoint with other components to a consistent software application. Some examples from the current
project developed for the Agency are used to illustrate the procedures described.

3.1 SharePoint as a Workflow Management System

In order to implement a workflow management system in a SharePoint application, lists, SharePoint work-
flows and event receivers have to be used. An entry in a list represents a process with the associated
metadata, such as the name, description, involved employees and event details.

Event receivers obtain the ability to inform other systems or applications about the creation or update of
a process. The source code of an event receiver has to include the necessary implementations.

SharePoint Workflows are used for creating tasks and distribute them to the users of a SharePoint appli-
cation. Several different workflows can be assigned to a list or document library. In addition, it is possible
to determine whether a workflow should be started automatically when a list item is created or whether
this is done manually by a user. Within the execution of a workflow further events can be triggered, list
items can be created or modified and data can be obtained from the users of the SharePoint application. If
hierarchies of processes are required, this can be realized with multiple lists. The association between pro-
cesses can be realized with key fields, similar to the concept of primary and foreign key in databases.

When entering metadata for a process, selection lists are often based on a pre-defined list of values.
There is a “PeopleSelector” field which can be used to select one or more users registered in a SharePoi
t application. For custom selection lists the previously described lookup fields can be used.

Apart from the possibility of entering metadata, SharePoint also offers a set of different views on the
metadata out of the box. In addition, it is also possible to extend, limit and create custom views.
3.2 SharePoint as a Document Management System

The requirements of a document management system can be implemented with SharePoint document sets and document libraries. Document libraries provide the possibility to save a document with associated metadata. Document sets allow sharing metadata across several documents. Usually documents have to be managed in hierarchical structures. Folders are a typical approach to achieve this. The use of folders is of course supported within document libraries and sets, but this approach results in performance issues for nested folder structures with many layers. A great workaround in this case is managed metadata. The folder structure can be stored in the central taxonomy storage. Entries
of lists or libraries will be tagged with one or more elements of the structure in the corresponding field for managed metadata.

Another feature of document libraries and sets is version control. Version control enables management and storage for multiple versions of the same document at the same time. A rollback to a previous document state is possible at any time.

Particularly interesting is also the close integration of Microsoft Office products like Word, Excel, Access, PowerPoint etc. in SharePoint 2010. For example, a group of employees can work on the same Word document simultaneously.

### 3.3 SharePoint as a Single Web Application for Both Systems

SharePoint obviously offers functionality for implementing a workflow / document management system. Equally important is that the two systems can interoperate with each other. A common task in this context requires creating or adding documents in relation to a particular process. Lookup fields are well-suited for this requirement. A lookup field has to be added to the fields of a document library. This lookup field holds a reference to the key field of the process list, for example a field with a unique process number. This way an entry of the document library can be associated with multiple items of the process list.

![Diagram](Figure2.png)

Figure 2
Association between lists with lookup fields

The association can be made visible to the end user by developing a web page with an element for entering the shared key value (usually a text box). In addition, the Web Parts representing the views of the document library and the process list have to be integrated into this web page. The views of the two Web parts then have to be filtered according to the entered key. Now only those entries are shown which include the key in the field which is determined by the view query.
3.4 Integration of SharePoint with domain applications

According to the preceding explanations SharePoint is suitable for the use as workflow / document management system in a single web application. Usually additional applications are required for specific tasks. Therefore integration with other applications is crucial for the suitability as part of a consistent software system.

In case of the Federal Environment Agency, SharePoint has to be integrated with an information system for chemical substances and their effects on the environment (ICS). The essential requirements for integration are:

- Store documents in SharePoint which have been generated in the ICS
- Notification of ICS about newly created processes and study documents in SharePoint and transmission of data that must be kept in sync

For the integration of both applications corresponding interfaces are needed. SharePoint already provides three different types of interfaces:

- Server-/ Client Object Model
- Web services (SOAP⁹)
- OData / REST Interface

The approach with the OData/ REST Interface was chosen for the following reasons:

---

⁹ “Simple Object Access Protocol” - [http://www.w3.org/TR/soap](http://www.w3.org/TR/soap)
• Access via the Server / Client Object Model requires the use of Microsoft’s .NET framework and therefore does not offer the necessary platform independence.
• Compared to the implementation of Web services, REST interfaces are lightweight and easier to implement in multiple heterogeneous domain applications.
• ICS uses RESTful Spring Services.

For implementing the OData / REST Interface Many frameworks and SDKs for various programming languages already exist. Since the ICS is implemented as a Java application, the Restlet\(^{10}\) framework with OData extension is used supportive\(^{[Sinn?]}.\) Using this framework, it is possible to create entity classes for all SharePoint lists and document libraries. Accessing lists and their entries is encapsulated by methods which initiate the sending of HTTP requests to the SharePoint REST service in the background. By now the upload of documents and creating, updating and deleting list items is supported by the service. There is one restriction on managed metadata: Fields of this type cannot be retrieved or set via the Odata/REST service. If there is a need for manipulating fields of this type or other unsupported functionality, an additional, separate service has to be implemented.

On the side of ICS Spring RESTful services are used to implement the interface. Therefore ICS also provides a REST interface which can be used out of the SharePoint application. Calls to this interface are initiated by event receivers. For example, when a new process is created, an event is triggered which is intercepted by the corresponding event receiver. Within the implementation of the receiver a call to the ICS REST interface is send.

4. Summary and Conclusion

The main advantage of the approach with SharePoint 2010 is the configurability of SharePoint applications. Any changes that do not require a separate implementation such as event receivers, custom Web Parts and code workflows can be done from the web interface or the SharePoint Designer by authorized users. The following list illustrates an exemplary selection of possible changes:
• Create, modify and delete lists, libraries and their fields
• Create, modify and delete views on lists and libraries
• Full control of the user management
• Customize the look and feel of SharePoint web sites and also the functionality using Web Parts
• Create, modify, assign and delete configurable workflows
• Customize the SharePoint Search
• Use external data sources within SharePoint

Thus SharePoint users do not need to hire a software developer for implementing change requests. Another advantage of the configurability is justified in the simplicity of expansion and maintenance of a SharePoint solution. In addition the new OData / REST interface in SharePoint 2010 simplifies the integration with different applications on different platforms. Gaps in the functionality of this interface can be closed by developing custom services. Because SharePoint is developed by a large software company, bugs and security issues are solved fairly rapidly.

However, some difficulties result from the fact that the challenging requirements must be mapped to the existing functionality of SharePoint. This applies at least if the previously mentioned advantage of configurability and the possibility of using a procedure with less custom implementations should be unaffected.

SharePoint is a comprehensive and versatile product which can lead to a large overhead of functionality in certain scenarios, although many SharePoint features can be modularly added or removed to a SharePoint application. This may affect the intuitive handling and performance of a SharePoint application.

\(^{10}\) Java Framework for creating RESTful Services - [http://www.restlet.org](http://www.restlet.org)
If these issues can be solved, SharePoint enables flexible and convenient integration of executive domain applications with workflow and document management as desired by the authorities.

Bibliography