Using SharePoint to manage and disseminate fusion project information: An ITER case study☆

Barry Prescott [a,*], James Downing [a,b], Marco Di Maio [a,c], John How [a]

[a] ITER Organization, CS 90 046, 13067 St. Paul lez Durance Cedex, France
[b] Tessella plc, 26 The Quadrant, Abingdon Science Park, Abingdon, Oxfordshire OX14 3YS, UK
[c] projectglobe.com, 108 Totteridge House, 15 Yelverton Road, London SW11 3QH, UK

ARTICLE INFO

ABSTRACT

The ITER Organization, in common with many other fusion laboratories, has an authenticated-access website devoted to the communication of information to all its staff and remote collaborators. In 2007 and 2008, the number of registered users of this site increased by more than a factor of ten, to over 3000 at present, and with approximately 900 unique users using the website per month. In parallel, the project management of the organisation has been put in place.

A decision was taken to move the web platform from simple HTML to Microsoft SharePoint [1] and to web-enable the many applications and databases used for ITER management. This decision has been well justified by the power and extensive flexibility provided by SharePoint, for example it permits different groups to publish their own information and to collaborate, and to consolidate disparate spreadsheet data in linked SharePoint lists to improve quality and maintainability.

This paper examines the use of SharePoint at ITER: why it was selected and what benefits it brings to both the local and remote ITER community. Some active case studies are presented. The paper also looks ahead at what future benefits to ITER this platform offers, and reviews the type of information that the site can profitably publish.

The paper also highlights some of the limitations of the platform, the problems of integration with other ITER systems, and discusses its potential for adaptability in other scientific organisations.

1. Introduction

ITER, like any organisation, has much of its knowledge contained in people, e-mails, documents, and spreadsheets. This presents a number of problems, such as:

- when people leave knowledge can be lost,
- information is hard to find, and
- duplication can lead to inconsistencies.

The Technical Website at ITER was set up to provide a solution to these problems; it is a central source of technical information on the project. Previously the site was a set of static HTML pages maintained by a few individuals. As ITER grew, it became apparent that a more flexible system was needed, with the ability to offer different security levels, to quickly update and create pages, to delegate the maintenance of parts of the web, and to collaborate without locking up information within e-mails.

At the same time, groups wanted to be able to manage structured data in open systems. IT was keen to avoid having a multitude of different systems, servers and platforms to maintain, and so SharePoint was selected as the new standard web platform for ITER.

After a year of use, it is possible to review the benefits and limitations of SharePoint at ITER, and the wider fusion community.

Firstly, the content management and collaboration features of SharePoint and their use at ITER are considered. These were the features used to replicate the information that was contained in the old Technical Website.

Next the paper examines the use of SharePoint's structured data storage (which is achieved by the use of data structures called 'lists'), and its applicability to scientific and engineering data.

Finally, the future of SharePoint use at ITER is discussed; the new features that ITER will exploit to provide a portal onto data in the many ITER data systems.

2. Content management

Most large organisations need to publish information on an intranet for internal purposes. The following examples show how the publishing functionality of SharePoint is used in the Technical Website at ITER:
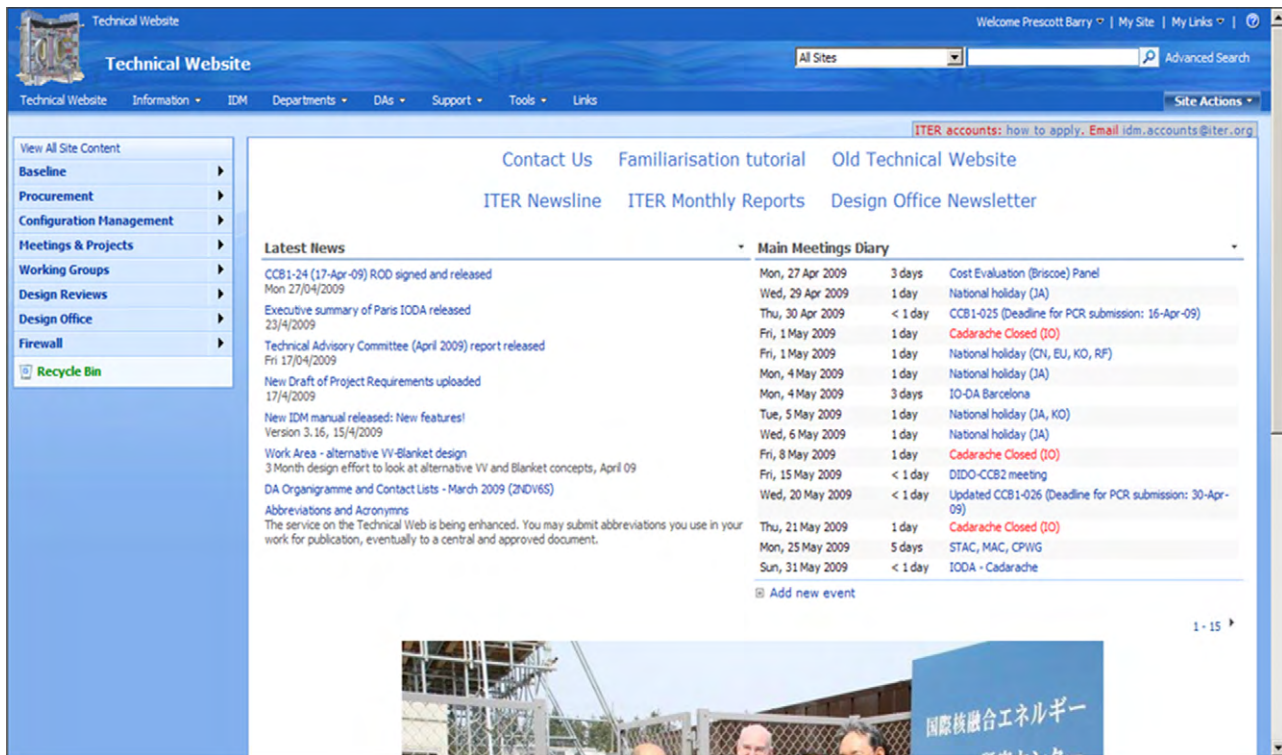
Fig. 1. Technical Web front page.

2.1. Project news and calendar

The cornerstone of the Technical Website is the front page where project news and updates to the site are regularly disseminated (see Fig. 1). Beneath the news is a regularly updated image of an important project-related event. To the right of the news is a window onto the project calendar.

2.2. Baseline documentation and project information

The historic Technical Website containing the past and present project baseline documentation (from project specification to design description documents) was imported into SharePoint. This demonstrates that static historic web content using relative links may be held within SharePoint without major change.

Web areas for Configuration Management, Project Change Requests, Working groups, Design reviews, Integrated Product Teams (IPTs) and Alternative Design options also all have a place on the Technical Website.

2.3. Abbreviations database

One particularly useful utility is the central list of abbreviations and acronyms, which tend to flourish on large projects. The abbreviations database on the Technical Website is based on a SharePoint list that is populated by user submissions which are moderated using the content approval feature. After a user submits an abbreviation and its definition, an administrator checks the validity and then approves the abbreviation so that it is viewable to all users of the website. Thus, coherency and standardisation in the use and description of abbreviations in the project is ensured. Moreover,

a report is generated regularly and approved within the ITER Electronic Document Management System (IDM).

2.4. Other information sources

The ITER library catalogue, book requests and on-line access to books and journals is managed from SharePoint. Useful administrative information, social pages, and a notice board are available for staff.

2.5. Departmental sites

ITER departments may publish information presenting their mission and progress, etc. Maintenance of the ten different departmental sites is delegated to trained site owners in those departments. This removes the bottleneck of every update to the web having to pass through a small number of webmasters to be effected.

2.6. Permissions and security

Access to the stored data must necessarily take into account the provisions of the ITER agreement as well as Quality Assurance, procurement, regulator needs, and legal requirements. As with IDM, the Website must ensure controlled access by ITER Domestic Agencies (DAs) and other selected entities to information which needs to be shared by ITER.

A harmonisation of the security groups available in the IDM and SharePoint, based on Active Directory [2] is planned. SharePoint's permission system is already based on Active Directory. Thus, someone publishing something who wishes to set access permissions on some content may use the same user groups whether he is publishing in IDM or in SharePoint.
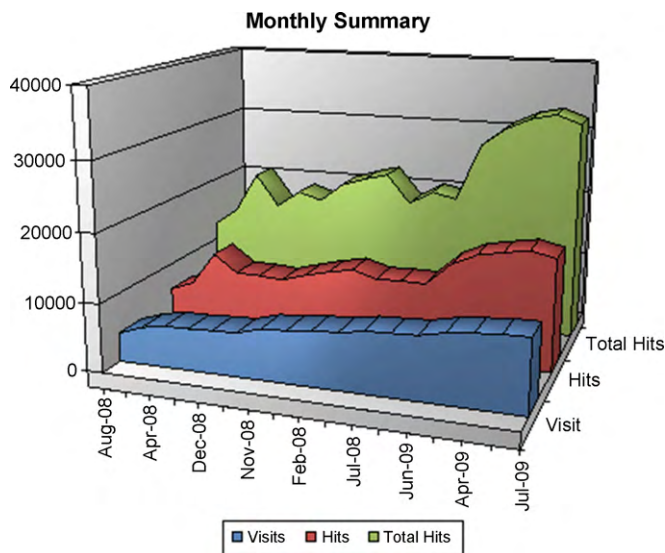
Monthly Summary

Fig. 2. Average daily requests by month (includes weekend data). Note the lower usage periods in August and December (holiday seasons).

2.7. Website usage

Fig. 2 shows that the usage is increasing steadily with time.[1] In the near future, a third party reporting tool will be added to perform better usage analysis than the standard SharePoint tools allow. By analysing where users go and what they search for we intend to improve the contents to increase the utility of the website to ITER.

2.8. Limitations

SharePoint has a learning curve which is initially steep for content authors. In order for authors to do their jobs effectively they need to get to grips with many concepts such as Sites, Web Pages, Document Libraries, Folders within Document Libraries, Content Types, Lists, Permissions, SharePoint groups, Active Directory security groups, Web Parts, Placeholders, and Navigation. It is more difficult to manage images than text. They may need to learn HTML and XSL to present the data as they wish. After that, there is workflow, templates, audiences, and search to consider.

Fortunately, as SharePoint is a popular product, there are lots of sources of information and training courses available. Once learned, the technology allows users to present information and data in a flexible manner, with less effort than a more basic technology such as static HTML. The return on investment of learning how to use SharePoint publishing is high.

3. Collaboration

A large project faces many collaboration challenges. Bates and Smith [3] identify these as:

- difficulty managing project-related information
  Information is spread and hidden throughout the organisation, duplication is rife, significant time is spent searching for, or helping someone search for up-to-date information.
- lack of communication

---

[1] Legend of Fig. 2: Visits: if a user browses to the Web site and views several files, it is all counted as one visit. Hits: Web pages only. Hits to supporting files, such as jpg files, are not counted. Total Hits: includes hits to all files including supporting files.

Decisions, scope changes, and updates to objectives may not be communicated to involved people, leading to redundant work, missed deadlines, and frustration.
- coordination of geographically dispersed teams
  Project teams are created based on skills and costs rather than geographic proximity. Project materials need to be accessible despite geography and time zone.
- lack of visibility into the project
  All project stakeholders need to be able to review project materials, understand the status, and be made aware of changes to the scope and objectives.
- managing changing resources
  Over time, people's availability for a project will change. Knowledge transfer is critical.

SharePoint helps ITER users to manage project-related information over the network in a more structured and useful way than using individual local copies of information and e-mail, or even using shared folders. It does this by providing a central repository for project materials that is accessible from anywhere connected to the Internet. For document management, ITER uses IDM rather than SharePoint; but for data management, simple project management, and communications SharePoint is often used, resulting in reduced time for users to find relevant information.

3.1. Group sites

Communication to a group can be improved by the use of SharePoint collaboration sites which provide facilities such as the ability to publish announcements, to share calendars, and to track the status of tasks. At ITER, sub-departmental groups (divisions and sections) and cross-departmental teams such IPTs are provided with collaboration sites to permit the groups to structure their work. Designated site owners may maintain their sites and add pages of information and sub-sites such as wikis and meeting workspaces. The CODAC group at ITER use SharePoint extensively to collaborate. It is the example used in the rest of this section.

When a new member joins the CODAC group, the SharePoint site owner requests that the new member subscribes to the team's announcement list via SharePoint's e-mail alerts or RSS (Really Simple Syndication).

3.2. Collaborative work on documents

Users can create a document workspace to collaborate with others at a central location on a document. The integration between SharePoint and Microsoft Office offers useful features for collaborating on documents such as workflow, check-in/check-out, and versioning. When the document reaches a stage where it is ready for publication outside the circle of its authors and initial reviewers, it can be uploaded to IDM where the formal document approval processes are enacted.

The CODAC group most often do collaborative work on documents thus: An author who would like feedback on a document uploads it to the shared documents folder on the site, and announces it in the announcements list. As everyone in the group is subscribed to announcements, they are all notified of the existence of the document. People check out the document, modify and/or comment on it, and then check it back in. While the document is checked out by someone, no one else may work on it. Intuitively this seems limiting but experience has shown this is not a problem in practice. Working this way the author has no need to expend effort collating comments.

ITER uses a system of three letter codes to represent people. Use of these codes has good acceptance in the CODAC team who report that they work if the team is not too big and people are able

to remember all of the codes. The codes are used in collaborative editing, i.e. everyone has set up their initials field in MS-Word so their comments and corrections in documents appear under that code.

CODAC sometimes collect a lot of comments from outside the section in large MS-Excel tables. They collate all the Excel files together and import the comments in a SharePoint list. SharePoint permits item-level locking. There is thus no need to lock the whole Excel file during the editing process. When the editing is finished, the comments are serialised back into Excel. This method is very convenient when many people need to work on the same document at once.

3.3. Shared calendar

The ITER CODAC calendar system was tuned by the CODAC group to suit their needs. It contains conferences, missions, vacations, meetings, and other types of events. The CODAC secretary adds in the calendar entries from SAP, mission forms, etc. Some people also do this for themselves. The site owner does the rest from the information he gets during the team progress meetings. The data captured can then be used to facilitate planning. For example, a list of conferences and attendees is automatically generated. The disadvantage of double-entry of leave into SAP and SharePoint could be addressed by interfacing the two systems, and this could be worth doing if such shared calendars will be deployed Organisation-wide.

3.4. Meetings

Meeting workspaces provide a place for meeting participants to share objectives, an agenda, relevant documents (or links to them) and resultant minutes. By using SharePoint meeting workspaces to draw up an agenda quickly, invite participants, record attendees, note decisions, and attribute tasks, one author of this paper has found that it is no longer necessary to record the minutes of meeting in a document. Much time is saved and "political" problems can be avoided by not recording discussions, but only decisions. The resultant SharePoint data has executable behaviour which facilitates the successful completion of the actions decided upon in the meeting.

3.5. Task tracking

The CODAC team manages its tasks in SharePoint. They have a meeting workspace for their regular team progress meeting. Actions are entered in the meeting workspace and are later copied to the site task list which is also shown on the CODAC front page which is used for tracking. This method is found to be convenient, but it requires having a person to oversee the whole process.

3.6. Sub-sites

• Task-specific sub-sites

The site owner may create a special purpose sub-site and let a user administer it. The site may be destroyed once its purpose has been achieved. Deleted items are stored by SharePoint in a recycle bin from which the user himself may retrieve items. This facility alone is a huge improvement over shared folders which do not facilitate the restoration of a deleted item by the user.

• User blogs and team wiki pages

People have their own blogs to share information with others. CODAC use blogs to communicate and record for future reference a more developed description of a topic that might have been discussed informally.

3.7. Limitations

The principal limitation of collaboration sites is the manpower required to educate users about the facilities available and then motivating and training users in their use. Once users appreciate the utility of the collaborative facilities of SharePoint, they are generally keen to use them.

4. Data management

ITER has produced and will continue to create an enormous amount of information that can originate from anywhere within the organisation. Its value to the project depends on factors such as its quality (to what level is the data correct, appropriate, and up-to-date?) and its ready availability. Using SharePoint lists has raised the awareness of data management amongst users and helped to consolidate information and significantly increase its quality and availability.

4.1. Scientific data management

ITER data requirements are complex: scientists, engineers, and other staff need to be able to share, connect, and analyse data throughout the extended organisation, preferably in real-time. The highly heterogeneous mix of skills and expertise leads to a multitude of software solutions employed and information formats used. Connecting the individual data sets is almost impossible and a culture of data standards has not yet been established.

Traditionally information and data have been managed in the form of documents and spreadsheets. The main advantage is that everyone can create documents and spreadsheets in layouts suitable to their current needs, which gives rise to a sense of ownership. However, this approach has severe implications in large scale projects because multiple independent authors tend to create incompatible layouts and formats. Not only are incoherent spreadsheet layouts difficult to read by humans; they can be impossible to process by machines and expensive to maintain; the need for a more controlled environment has become a priority on the ITER road-map.

4.2. Local data management

In the absence of an ITER-wide data system, a practical solution had to be provided. A pilot project was conducted in the Diagnostic Division (DIAG). Their data was available in a number of spreadsheets, all designed to cater for a specific need, using and re-using data both controlled by DIAG and referenced from other divisions. The objective was to qualify, consolidate and analyse the existing data and migrate it to a new dedicated SharePoint site where it could be managed centrally and controlled by workflows. Providing the data on the server also made it available to all interested parties, significantly reducing the external requests for DIAG data which traditionally resulted in manually 'customised' MS-Excel files sent out by e-mail to the requestor.

The existing spreadsheets were identified, frozen, and analysed. Most showed a matrix-layout, typically listing a number of properties over individual diagnostic systems grouped by, for example, the party providing the system, the Responsible Officer (RO) or the location in the machine. Owing to this, information was redundant, data formats were non-coherent, and many cells were empty (see Fig. 3, left hand side).
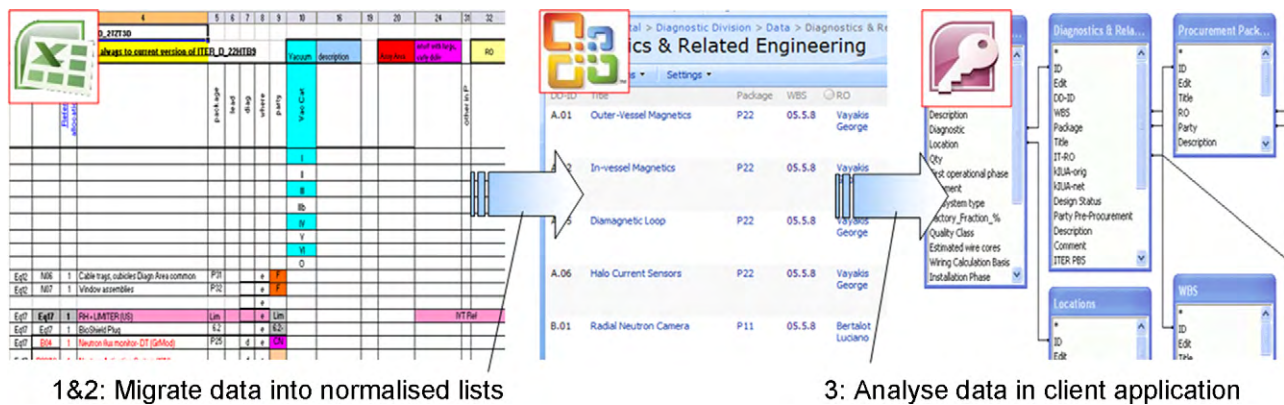
Fig. 3. The process of migrating to SharePoint: (1) create normalised list structure in SharePoint; (2) migrate data in from Excel using ActiveX controls; and (3) link data back to Excel and Access to generate reports, pivot tables, and other analysis as required.

Fig. 4. SharePoint list 'Diagnostics'. Note that nearly every cell is a link to more information.

In a first step, the data was analysed. The individual data structures were identified and broken down into their constituent objects. Each of these was collected in a Custom List on the new site and linked to other lists on that site (see blue cell entries in 'Package' column in Fig. 4, which directly open the package information related to that item).

It was found that using the "Import from Excel" feature to create the lists often did not lead to the expected results: the layouts and columns created did not always match the data format required. Creating lists manually in SharePoint and then importing the data (either by connecting the new list back to Excel or by using 'copy and paste' functionality in the Datasheet mode) became the standard method of migration. Acceptance of the new centralised data system was low to begin with. Unfamiliarity with the concepts of SharePoint led to a number of iterations to perfect the data structures. The high density of information, however, was immediately apparent: almost every selected cell contains a hyperlink to more data. After a critical mass of information was reached, the server data was defined as the only master copy, and versioning on the lists was activated then DIAG accepted and used the system. The main advantage of a centrally controlled data server, namely to offer better control over the quality of the data, only became obvious after connecting the lists in MS-Access and training key users to create custom analyses and reports (see Fig. 3). This allowed them to re-create the 'original' spreadsheets from up-to-date on-line data, although they were not as colourful.

4.3. Limitations

There are several limitations to using SharePoint as a data storage solution, some of which are technical and inherent to the SharePoint data structure, while others were due, largely, to time limitation, and could be overcome with more time and experience.

The major technical limitations are that:

- lookups are limited to displaying one column at a time,
- list performance degrades with list view size,[2] and
- it is difficult to connect list data between sites.[3]

These led to work-arounds: replicating data from other divisions without any automatic synchronisation (much of that information is still mastered in MS-Excel sheets!). ITER is currently addressing the issue by replacing the copied information with live-links using the Business Data Catalogue feature of SharePoint and custom webservices.

Another limitation found was the lack of in-built server based analysis and reporting tools. Without MS-Access and training, it is

---

[2] This is a limitation of the standard view components when the list view contains more than 2000 items.

[3] This would be possible with custom development, but not by using out-of-the-box features.
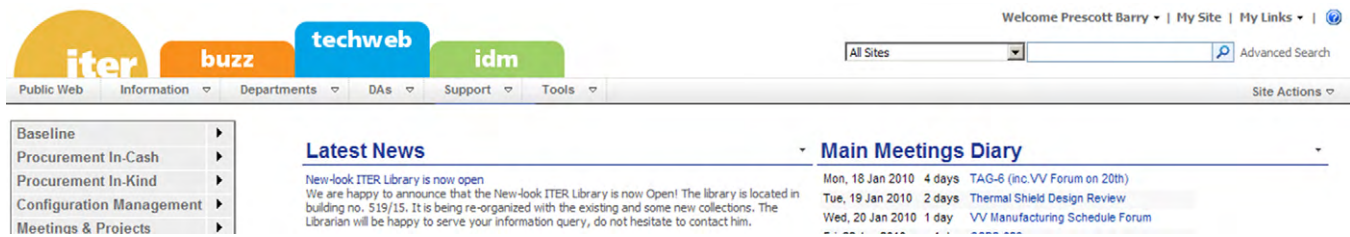
Fig. 5. Mockup of tabbed-based navigation to integrate all ITER websites.

impossible to re-create the former spreadsheet layouts. Third-party components have been evaluated, but so far, nothing powerful enough to provide MS-Access type features has been found. In addition, no satisfactory solution for simple graph representations of list data could be provided.

Other minor problems have been encountered, but they could probably be overcome with more effort. For example, it is currently not possible to:

- perform adequate analysis on columns which contain multiple values,
- define the owner of an individual list item (for example, the measurement table contains information from many individual responsible officers (ROs), unless the RO is also the creator of that item), and
- provide a simple means for defining more complex, custom workflows.

4.4. Outlook

The DIAG pilot project has been devised as a temporary, practical solution to get the data out of the spreadsheets and into a database-like structure for future migration. This has greatly improved the efficiency of providing high quality data to their organisational interfaces as it allowed the controlled collection of data from multiple sources and its distribution back into the community on-demand. It has also raised the awareness on data quality and control through the introduction of approval workflows thus reducing the amount of manual work and cost of maintenance.

The division is almost self-sufficient and is widely regarded as a successful model for ITER. Following their example, other divisions have set up similar sites. Once an ITER Product Data Management (PDM) system is available for everyone to manage their data, the now familiar SharePoint could be used as the front end to that data, at least in simple, read-only scenarios. Its data storage capabilities may still be used, but limited to locally stored data.

5. Using SharePoint as a fusion portal

The previous sections of this paper have discussed the features in SharePoint that ITER is currently utilising; there are many features that ITER is not yet exploiting. This section looks at the direction that ITER hopes to take SharePoint, specifically with regards to expanding its role as a portal for fusion data.

ITER has many different systems to handle different types of data, for example: Primavera [4] for scheduling data, SAP [5] for financial data, and Enovia [6] for CAD drawing management. Each of these systems exists in isolation from the others, requiring a different application to view and edit the data. There is, however, a need to view the data from these systems in a single place [7]. This would not only enable users to find the information faster (they only have one place to look) but also allow them to view related information from the separate systems in one place. In the above example, a user may want to view the schedule related to a part, its current design, and it's estimated vs. actual costs without having to locate it in three disparate systems.

The out-of-the-box features, along with limitations, and the programmatic customisation options that are possible are considered below for three specific paths for enhancing the portal.

5.1. Navigation

The first improvement considered is the navigation across the different applications. Currently a navigation bar is shared between the two main web applications at ITER: SharePoint and ICP (an in-house data management system). While this allows consistent navigation between the applications, it has lead to confusion over which pages live in which application. As more features, sites, and applications are added, the menu will become unmanageable.

A tab based menu (or similar visual metaphor) will be employed (see Fig. 5), allowing tabs to be allocated to different applications, improving the usability [8]. This will clarify to users which application they are currently in, and let them launch any web application from wherever they are.

However, with tab-based navigation, the data will still live in separate applications, and so by this method alone it will still be impossible to view data from different systems on the same page.

5.2. Search

Although changing the navigation will help users navigate to content, frequently they will not know exactly where to look, or what other similar information is available; forcing them to search for it. Most of the applications at ITER have some search functionality built in, but again the user is forced to visit each application in turn to search them. As an example of the confusion this causes, a number of zero-result queries in SharePoint are seen every day for identifiers of documents stored in IDM.

The SharePoint platform has advanced search capabilities [9]. It can index multiple sources of content, and return unified results, or results broken down by various categorisations. There is also a community-supported 'Faceted Search' component that allows users to narrow down a results' set by suggesting relevant filters to apply. Other products have similar capabilities, but as SharePoint has been chosen as the portal for ITER, it is logical to use it.

The four features of SharePoint Search that could be used for ITER's needs are:

1. federated search,
2. crawl as a website,
3. the Business Data Catalogue (BDC), and
4. writing a custom Protocol Handler.

Each of the above has its own advantages and disadvantages, and so each will be suited to different applications.

1. Federated search allows SharePoint to display results from another search engine. This is good for systems that have their
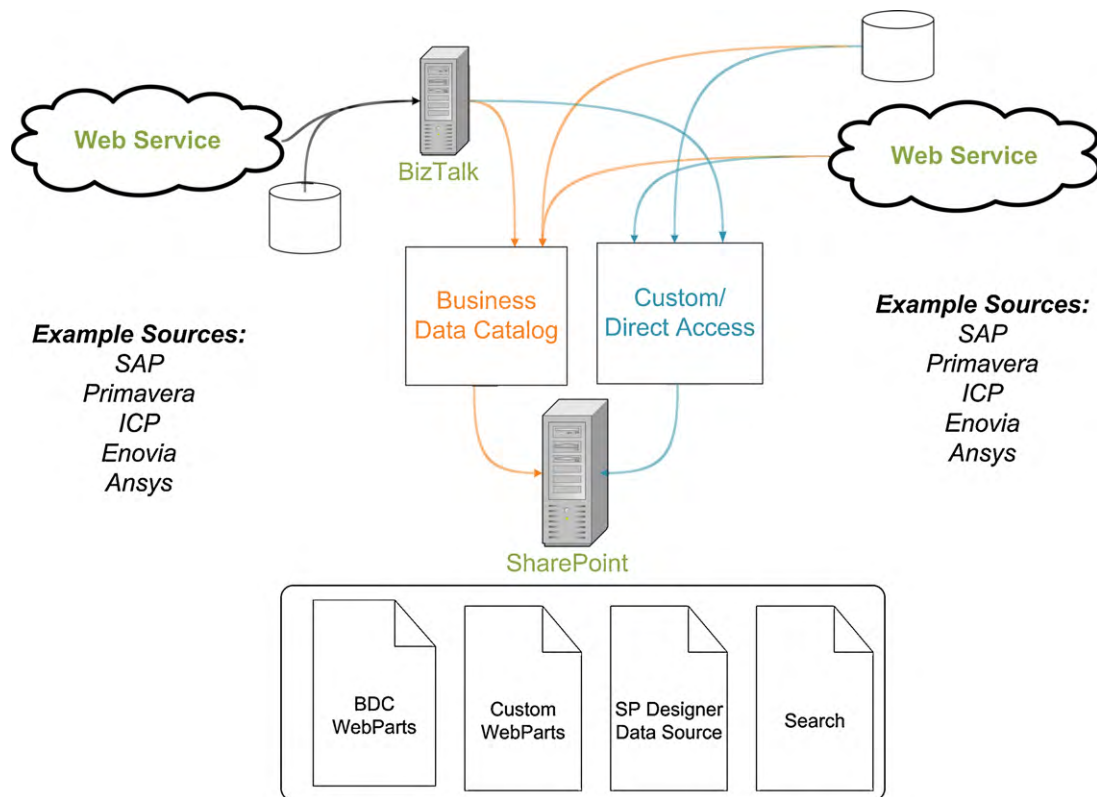
Fig. 6. Architecture for gathering data from external sources.

own search engines and where results do not have to be ranked with results from other systems. The other system can then handle the security trimming, and any specifics about the ranking of the results. For example, search engines hosted at ITER DA's could be used in this way.

2. Systems with web interfaces can be crawled as websites. This is quick to set up in SharePoint, as virtually no information needs to be provided about the application other than its URL. However, structured data cannot be recovered from the pages, and no security information can be collected using this protocol. For some information sources, this does not matter; for example, this method is used to index ITER's public website, so that it is searchable from the Technical Website.

3. For systems with structured information, the BDC offers excellent out-of-the-box functionality. The BDC definition file instructs SharePoint on the structure of content in the application and how to retrieve the content to index. For example, this can be used to index Work Breakdown Structure (WBS) scope descriptions from Primavera. There are some limitations, however, as SharePoint cannot index certain field types from the BDC (e.g. binary data) and security lookups occur at query-time, not index-time, which can dramatically slow down queries.

4. Finally, it is possible to write a custom Protocol Handler for SharePoint; permitting the greatest flexibility. This comes at the cost of developing and maintaining a custom solution. The protocol handler can pass back a mixture of structured and unstructured data from an application, along with security access lists, through a well defined API.

5.3. Composite views

The final goal of SharePoint at ITER is to be able to display information from different applications on a single page. SharePoint is ideally suited to this as it automates the process of creating sites and pages for users. Once a user has a site, he can easily create and link components on pages to display exactly the information in which he have an interest. The components can be security trimmed.

SharePoint provides two mechanisms to transmit a user's security credentials. Kerberos provides a secure way to pass the credentials from one system to another, as long as both systems support Kerberos authentication and trust one another. The second option is the SharePoint Single Sign On service (SSO). This stores user credentials for sub-systems, so the user only has to enter them once.

ITER is implementing BizTalk [10] as the messaging server for integrating applications. This can be integrated with SharePoint either via the BDC, which automatically generates pages and reusable components to display and connect the data, or via custom written components which may be more flexible but would require maintenance. Fig. 6 summarises the architecture for this integration, showing how external sources can be configured to connect to SharePoint and have their data displayed in various user configurable components.

The first systems to be integrated in this way will be to show the Work Breakdown Structure (WBS) baseline information and selected documents from IDM in SharePoint in a non-folder based view.

6. Conclusion

SharePoint has had a beneficial impact at ITER, providing a unified web platform for users who want to create and maintain collections of web pages. These ad-hoc pages provide a user friendly way to browse the project's fusion (and non-fusion) information. We have observed high usage of the workspace and collaboration features, for organising meetings, drafting documents, and communicating ideas within the ITER Organization and with members of the DAs.

The benefits above stem from the fact that a fusion organisation, especially one distributed across multiple continents, like ITER, has to deal with many of the same problems as any other large organisation: stakeholders have to be able to collaborate and communicate effectively. However, a fusion organisation also has to deal with large volumes of scientific and engineering data. It was found that while the interface for data management in SharePoint provides a simple way for people to extract their data from documents and store it in a relational fashion, the engine behind SharePoint has limitations for large shared datasets, and for performing complex aggregations on the data. SharePoint has nevertheless been instrumental as a bridge into a dedicated engineering database.

For the future of SharePoint in fusion, increased use of Share-Point as a portal onto other data is foreseen, with the ability to search for and find data in other systems, and display up-to-date data from multiple systems on a single page, through integration with BizTalk and the Business Data Catalogue.

The main difference between ITER and the commercial companies at which SharePoint is targeted is the complex scientific data produced. While SharePoint is weak in managing this data, its strengths lie in the collaboration and portal components fitting around such data. These features benefit any large project, including scientific ones. For this reason, we think that other scientific organisations would do well to consider this platform.

References

[1] SharePoint, URL: http://sharepoint.microsoft.com.
[2] Active Directory, URL: http://www.microsoft.com/windowsserver2003/technologies/directory/activedirectory/.
[3] S. Bates, T. Smith, SharePoint 2007 User's Guide, Apress, Berkeley, CA, 2007.
[4] Primavera, URL: http://www.oracle.com/primavera/.
[5] SAP, URL: http://www.sap.com.
[6] Enovia, URL: http://www.3ds.com/products/enovia/.
[7] J. Firestone, Enterprise Information Portals and Knowledge Management, Butterworth–Heinemann, US, 2003, pp. 35–42.
[8] S. Krug, Don't make me think!, New Riders, US, 2006, pp. 79–80.
[9] B. English, Microsoft Office SharePoint Server 2007 Administrator's Companion, Microsoft Press, Redmond, 2007, pp. 555–556.
[10] Biztalk, URL: http://www.microsoft.com/biztalk/en/us/.