Microsoft Response to the government's proposal
Sharing and collaborating with government documents

Summary

Microsoft believes that the government will only meet its objectives of reaching the most people at the least cost for all if it includes the most popular and most widely used open standard document format – Open XML (ISO/IEC 29500) – in its short list of standard formats for sharing and collaborating with government documents.

The inclusion of HTML in the government's proposal (alongside ODF) means the government recognises the value of avoiding a single standard for documents, but using HTML for discrete documents (or for more complex content like that often found in a spreadsheet) which can be edited "off line" is not practical, making the sole choice in the government's proposal effectively ODF.

Mandating one open standard for discrete document formats over another completely ignores benefits enabled by a choice of modern formats and is therefore likely to increase (not decrease) costs (*link to original open standards consultation*), risk widespread citizen dissatisfaction (which the government is attempting to avoid) and add (not remove) complexity to the process of dealing with government

Microsoft believes that the least cost and most effective way forward for any organisation seeking to ensure the maximum range of interoperability, the richest range of functionality and the widest use of common formats should be to embrace multiple open standard document formats e.g. both OpenXML (ECMA-376, ISO/IEC 29500-1:2012) and ODF (OASIS ODF v1.1, ISO/IEC 26300-1:2006).

In this response we set out to address each element of the proposal and to present evidence that shows the government risks increasing costs and reducing interoperability by ignoring the fact that the vast majority of citizens and businesses already use OpenXML as their preferred document format. While including ODF is a choice that Microsoft supports, ignoring and omitting OpenXML will ensure that the very things the government is trying to avoid are actually more likely to happen.

In this response

We will show that OpenXML enjoys a popularity of use across the major domains of interest to the government, including across the major domains of the UK Public Sector (gov.uk and nhs.uk) and of UK business (co.uk) and of the 3rd sector (.org.uk).

We will also show that both standards (OpenXML and ODF) are vibrant and developing, that they are both supported by a range of tools (applications, apps, programs or services) widely available to the government, to business and to citizens in a range of circumstances to suite user needs.

And we will also show that it is the nature of standards to develop and occasionally supersede one another, and that for a time, multiple standards that appear to do the same thing can and do exist and thrive in parallel for many years to the benefit of all.

On this basis, we urge the government to avoid a costly and unnecessary focus on too narrow a selection of standards. Having already recognised in its proposal that no single standard will be

adequate for its needs, we now urge the government to include OpenXML alongside ODF and HTML in its list of standards for sharing and collaborating with government documents.

Addressing the proposal
In the sections below we have used the convention of repeating statements from the government's proposal in *bold italics* and presenting our response in normal text in the subsequent indented paragraphs.

Citizens, businesses and delivery partners, such as charities and voluntary groups, need to be able to interact with government officials, sharing and editing documents.

> Citizens, business and delivery partners, such as charities and volunteer groups are able to interact best with government officials, sharing and editing documents if the latter embraces the widest range of international open standards for document formats, a choice that should include both OpenXML (OOXML) and Open Document Format (ODF). Evidence of use already shows that most documents that can be found on the internet in domains most closely associated with these groups use OpenXML over one hundred times more frequently than they use ODF.

Officials within government departments also need to work efficiently, sharing and collaborating with documents.

> Officials within government can work more efficiently, sharing and collaborating with documents if they embrace any open standard format other than Microsoft's old Binary formats. Both OpenXML and ODF are supported by a wide range of productivity tools, including all version of Microsoft Office since 2007. OpenXML and ODF both offer dramatic reductions in file size, being based on data compression technologies. Officials already enjoy considerable familiarity with Microsoft Office and the main benefits of efficiency come from having access to the most effective and flexible tools, not from the format that any output is eventually saved in.

Users must not have costs imposed upon them due to the format in which editable government information is shared or requested.

> By embracing and publicising that the government can create and consume content in a range of common open standard formats, the government will avoid imposing undue costs on any segment of society. Making too narrow a choice (e.g. only ODF) risks alienating those citizens and businesses that either have no capability to read and edit ODF or who have already chosen to use OpenXML as their default. The capability is found lacking in tools included with products like Apple's iPad (which needs additional software to read ODF) or in services like Google Docs (which has recently deprecated its ODF support). By embracing both OpenXML and ODF, the government can guarantee to reach the widest audience across both citizens and businesses without obliging either to expend money just to communicate with government.

As technology progresses, government's production of editable information in formats traditionally associated with documents will become less important for users.

It is, perhaps, too soon to call an end to information shared in discrete digital objects (we can call these "files") in any format or another. On the contrary, easily found evidence shows that the creation and consumption of discrete digital objects (files) is increasing and that the increase is faster than at any time in the past. What is important is that the format for a file is based on an open standard that allows the content to be easily consumed, with or without original layout "metadata", either by a person or - increasingly important - a service. It is true that much more information published by any creator will default to one of a number of formats common on the Internet, e.g. HTML, which is an important standard to use in this context, although this should not be the only standard used. It is also true that information sought by one party (e.g. the government) from another (e.g. a citizen or business) will be provided through some sort of web site or web service. Increasingly, this may become a communication from one service to another (and back again).   In that context it will be important that the format of the data is identifiable between the services and that the format can cope with the richness of the content in an efficient way. But on either side of the transaction there will be, for a long time yet, a need to "take a copy" of the content in a format that can be stored, viewed and managed as a discrete digital object - a file. People will choose the format that most suits them. Government should seek to embrace the most popular, the most stable and the most developed standards to reach the most citizens and business.

Government services are being redesigned to make them more straightforward and easier to use by making them digital by default. This will diminish the use of traditional government document formatting even further as information is published or collected directly on the web.

We do not dispute this, in fact, we address this with the observations above. This is more a statement of direction than a requirement or evidence of a need supporting the proposal made.

Users need to open, edit and save information online and offline

This is possible in OpenXML as well as ODF and HTML.

Users need to submit information in response to a request, to perform a transaction or to access a service

This is possible in OpenXML as well as ODF and HTML.

User need to share information with specific people

This is possible in OpenXML as well as ODF and HTML.

Users need to publish information online so that a wide audience can access and work with it

Although a document saved as HTML may appear to have an advantage for online publication, in practice an element of re-formatting will always be needed because there are significant layout differences between a "web page" and an "A4 page". Collaboration during

the creation phase of information destined for publication can rely on one format, while the tools used will then readily convert the "collaboration" format into the "presentation" format (of, perhaps, HTML). Since these are widely supported open standards, there exists a wide range of tools able to do this, hence, this is more a requirement on the capability of the tools chosen in any given context than of the formats used at each stage in the document production.

Users need to edit information and be confident that it remains usable and editable when saved and shared with other users

There are two elements essential to the success of this requirement - the first is to use a mature, feature rich and widely supported format - OpenXML and ODF are both excellent examples and the second is the proven interoperability of the tools chosen to transact the documents. Interoperability between tools is a function of two things: [1] the capabilities of the developers of the tools and [2] the ease of interpretation of the detailed features of the relevant standards and how easy or difficult they are to implement. It is evident from the widespread support of both OpenXML and ODF that either standard is described in sufficient detail, with sufficient richness of features, for them to be successfully implemented by more than one party. Thereafter, what we see is that the subtle nuances of implementation, and how those implementations handle different content captured within different formats, become the subject of extensive interoperability testing. A "reference implementation" may be chosen and different data (content) is encoded in the standard format by the reference implementation and experts compare other implementations of the standard against it. This is a dynamic process and, to some extent, an on-going process for as long as the standard is in development. All the standards considered here are "in development" and, as such, there will always be minor differences of implementation to be resolved through interoperability testing. The point we are making is that this is not an absolute science and there is no automatic or guaranteed "perfect" implementation. Confidence is therefore built on the basis of experience and commodity. The available evidence suggests that the most common "modern" open format used is OpenXML and on that basis it is reasonable to expect that experience with that format is more likely to have resolved interoperability differences than with other formats. That said, it is important to recognise that there are many events organised to test the interoperability of products that use the ODF format [insert references to the Maidenhead ODF "plugfest" and the Brussels one). It would be an academic exercise to differentiate between instances of the implementation of either OpenXML or ODF and, for all practical purposes, their implementations should be considered equivalent (wherever they have been implemented and tested against their "reference").

Users need to create a new document with the same style as documents previously

This is, again, more a function of the capability of the editor or tool used to create the document than the format itself. While the standard for the format may allow certain features to be expressed in the layout ("style"), it is entirely in the hands of the developer of the tool as to whether their tool (application, app, programme or service) will implement all the capability of the standard. Hence, a user with access to the same tool will be able to create multiple documents in the same style (but with different content), whereas the

government cannot guarantee that a document sent in a given "style" sent to another user (which they may display, print or even edit the content in-situ) can be used as a template for other documents with different content. That is a capability of the tools used and not the format of the document.

Users need to export the documents created in a non editable format so that they can share a document as they intend it to be presented

This is a requirement of the tools used to create and manage documents, not on format in which the document exists in its editable form. This therefore sets a procurement requirement which then relies on the standard selected for viewing government documents.

Users need to export the documents they create in a format compatible with other software so that other people can use the information

This could mean the tool has the ability to save other or many formats or it could mean they are attempting to select the format that's most widely implemented, but could also refer to an interoperability requirement which we have already addressed above.

Users need to share information so that they can gather feedback

This is possible in OpenXML as well as ODF and HTML.

Users need to share information so that they can respond to a request for information

This is possible in OpenXML as well as ODF and HTML.

Users need to view/edit the information shared with them so that they can read/act upon the content

This is possible in OpenXML as well as ODF and HTML.

Users need to provide input on information created by someone else

This is possible in OpenXML as well as ODF and HTML.

Users need to copy and paste content from one source to another so that they can quickly collate pieces of information in one place

This is possible in OpenXML as well as ODF and HTML.

Users need to edit information created by an integrated system they work with so that they can add additional information

This is possible in OpenXML as well as ODF and HTML.

Users need to gather feedback on information they have drafted so that they can apply other people's recommendations to the content

> This is possible in OpenXML as well as ODF as revision marking is possible in both formats, but is only available depend on the choices made by the developer of the tool and in HTML only if the web developer builds such capability into their web site, but there is no standard mechanism in the HTML standard to capture such feedback - it becomes a unique feature of the web site.

Users need to see version updates so that they can be sure they're working on the latest version of a document

> This is possible in OpenXML as well as ODF where each standard contains the ability to record such data in the metadata of the file. In HTML this would be a unique feature of the relevant web site, since the HTML standard does not provide a prescribed way to record this.

Users need to access information from any appropriate place so that they can get on with their work.

> Success against this requirement hinges on the definition of "appropriate place". Even for HTML there may be places where it is impossible to render the content to be worked on appropriately to the task at hand. For example, just because a smart phone has a web browser compliant with the HTML standard does not mean that it is practical to use any and all web services that may allow for the editing of content necessary to complete any given task. A web browser does not specify the screen resolution or aspect ratio used (and even on a tablet or laptop, the user can re-size the browser window) and some browsers may attempt to "interpret" the content best for the screen size and resolution of the device it runs on. These are choices of the browser developer and beyond the knowledge of the creator or intended editor of the content. So what is an "appropriate place" in the context of the work needing to be done? One might ask whether this is more about an "appropriate device" that can be used in an "appropriate place" in the context of the work needing to be done. Again, there are no absolutes when satisfying this requirement. Some devices may be appropriate, but only in certain places and some devices may be used anywhere, but may only be appropriate for certain tasks (they can do some things, but not everything). What is appropriate will depend heavily on the choice of the user, the place, the time and the range of acceptable solutions to the task in hand. Both OpenXML and ODF find implementations on a broad range of devices, although the evidence suggests that OpenXML perhaps enjoys a marginally broader range of device support at this time (more web services support this and more phones, tablets and other devices support OpenXML), making the choice of ODF in this context a little less certain.

Users need their devices not to be clogged up with downloads

> This is a function of the tool (application, app or service), not of the format in which the data is then stored. There are apps that don't down load the file at all, but allow the user to work on the parts they can display. There are apps that download the file temporarily, permitting "cached" editing which is then saved to an on-line (cloud) storage service when the editing is complete (and the cache is deleted) and others allow the user to choose where to store the

file at any stage in the process. This is not determined by the format, but by developer choices for the tool.

Users need to ensure integrity of specific documents, e.g. audit trail for editing, versioning

Again, this is more a function of the tool than the format. A user may be able to edit a document on their smart phone and save it in any of the standards considered here, but they may not have the ability to view or change the meta-data associated with the file. Equally, they may edit the same file on their tablet and the tool they use there can view and edit the meta-data. The tool may even be able to remove all or most of the meta-data. The ability to control the right to edit the meta-data may also depend on the capability of the original author's tools (and not on which format it is chosen to write the data in).

Users need to use the information on the device and platform of my choice, for example laptop, tablet or smartphone

As discussed above, this is dependent on what "use" of the information the author of that information (content) may be expecting a recipient to make. It depends on the definition of "appropriate" for the place, the time and the device of the user's choosing. This does not imply that a user will be obliged to incur additional cost to use the information, but that the available uses are a function of the user's choice of tool, rather than the author's expectations, irrespective of the format of the content. In the most extreme terms, an author who requests recipient to inspect and edit data in a 200 column by 10,000 row spreadsheet is not, in any practical sense, allowing the recipient (the user) to perform that action (easily) on (say) a smart phone. Some choices depend on the scale and layout and complexity of the content and the user may therefore sometimes not have a "free" choice - but whatever choice they get, it is largely independent (or much less dependent) on the document format than on the content of the document and on the action that needs to be performed.

Users need to be able to use accessibility tools with information in online and offline formats

Accessibility is a rich and complex subject, dependent on the user's choice of device and, perhaps, the inherent tools (applications, apps, programmes, services) that they have available on that device. E.g. HTML does not, inherently, include any accessibility features in the standard, but the standard itself allows information to be displayed to be tagged so that applications enabling greater accessibility of the information can recognise and react to those tags. The same is true for other applications used to create content in other formats - accessibility is largely a function of the tool and more independent of the document format.

Users are able to efficiently share and work on editable government information

This is possible in OpenXML as well as ODF and HTML.

Users are not required to buy new software to submit or work with government information

There are many tools (applications, apps, programs or services) available at a range of costs to meet this requirement. An Internet search will reveal many capable of handling the major formats recommended here (OpenXML, ODF, HTML and PDF). Many devices now include software (inclusive in the cost of the device) that capable of handling these formats. Many citizens and businesses will already have software that is capable of handling these formats. Viewers which permit "copy & paste" of the content are available for no cost for all the formats proposed here, permitting (almost as a last resort) the user to import the content into a tool of their own choosing.

Users are able to re-use data and text, where licences permit

This is not a function of the document format. It is the responsibility of the data owner to set the intellectual property rights on the content within a document. None of the formats recommended here – OpenXML, ODF, HTML or PDF – assert or constrain rights over any content in any document created in the format.

Additional requirements (functional needs) given are that the format should support:

Characters associated with Unicode 6.2 for text based file formats (in accordance with the standards profile for cross-platform character encoding)

This is true of both OpenXML and ODF.

Digital continuity - having implementations that enable support for import of older formats

This is a function of the tools used, not a feature of the format selected (see above).

Use of metadata

This is a function of the tools used, not a feature of the format selected (see above).

Imports and exports to/from other applications

This is a function of implementation and interoperability between different implementations. Both OpenXML and ODF have many implementations. Many implementations are now an inherent (i.e. inclusive) feature of a device or service and one can easily argue that OpenXML has popular support over a broader range of popular devices and services. While this is true, we would still encourage the government to select both OpenXML and ODF for sharing and collaborating with government documents.

Fonts and graphics that are reusable in other formats

Graphics are another subject for a discussion of standardisation. For all practical purposes, both OpenXML, ODF and HTML can "embed" a range of common and popular standardised graphics formats within their documents.

Creation of templates

Both OpenXML and ODF standards permit the creation of templates for documents. HTML is more flexible and less prescribed in this context, having coding standards in which web templates may be stored, but with more freedom for the developer to define the "template" structure. In this context, that flexibility may not offer the least cost or most productive mechanism for sharing and collaborating with government documents (unless all recipients just happen to be web developers).

Citizens, businesses and delivery partners must be able to interact with government officials and services, or those working on behalf of government, sharing appropriately formatted, editable information.

This is more likely and possible, at lower cost to all parties if the government includes OpenXML along with ODF and HTML.

Users should be able to work on their device of choice and must not have costs imposed upon them due to the document format in which government information is provided or requested.
There is no better argument to support a wider range of document standards than this. Evidence shows that the most popular modern format in widespread use by citizens, businesses, charities and by government itself is OpenXML, making the choice of ODF alone quite perplexing. We urge the government to include OpenXML in its list of document format standards to ensure the broadest range of people have no costs imposed upon them.

Documents should be editable on different devices without loss of integrity - the information should not become spoiled.

This is not a function of the format chosen, but of the quality of the implementation of the format. No format is inherently easier to implement than any other in this space. Implementation extent is a choice of the developer, not the author of the document format standard or the author of any document in a particular format. Developers can and do choose to implement more or (often) less of the features made possible by the standard depending on the platform they are targeting with their tool (application, app, programme or service). Implementation quality is measured by collaborative interoperability testing - interoperability is not an inherent feature of any standard, it is a goal achieved through excellence in implementation.

When dealing with citizens, information should be digital by default and therefore should be published online. Browser-based editing is the preferred option for collaborating on published government information. HTML (4.01 or higher e.g. HTML5) is therefore the default format for browser-based editable text. Other document formats specified in this proposal - ODF 1.1 (or higher e.g. ODF 1.2), plain text (TXT) or comma separated values (CSV) - should be provided in addition. ODF includes filename extensions such as .odt for text, .ods for spreadsheets and .odp for presentations.

For statistical or numerical information, CSV is the required format, preferably with a preview provided in HTML (4.01 or higher e.g. HTML5).

Forms and information exchanges should be digital by default where this is enabled, therefore use of office formats should not be encouraged for the completion of forms.

For information being collaborated on between departments, browser-based editing is preferable but often not currently available. Therefore, information should be shared in ODF (version 1.1 or higher e.g. ODF 1.2). The default format for saving government documents must be one of the formats described in this proposal.

> There is no evidence in the proposal to support these recommendations. When evidence of use and implementation is gathered and examined, it suggests that the use of OpenXML is more prevalent than the use of ODF. This in turn justifies the claims made in this response and the assertion that the government will achieve its goals herein quicker, at lower cost and with a lower impact on citizens and businesses if it includes OpenXML and ODF and HTML in its list of standards for sharing and collaborating with government documents.

To avoid lock-in to a particular provider, it must be possible for documents being created or worked on in a cloud environment to be exported in at least one of the editable document formats proposed.

> This is again a requirement on the web services chosen. It is safer to embrace a slightly broader range of standards and enjoy a more widespread choice of services.

Information that is newly created or edited should be saved in one of the formats described in this proposal. There is no requirement to transfer existing information, unless it is newly requested by a user and shared.

> This is a matter of policy, but it may be unavoidable that some documents have to be converted. Absolute prohibition of this conversion will be difficult and impractical to manage, monitor or control. Like it or not, any new standard (mandated or simply recommended, since, following RFC2119, the use of "should" above implies a recommendation, not a mandate) will drive behaviors that are unnecessary or undesirable. The simple inclusion of OpenXML in this standard will obviate the need for much of this unnecessary conversion.

A government body must not refuse to accept or supply a document in at least one of the open formats described in this proposal

> This is a statement which seems to imply that a government body MAY refuse to accept or supply a document in any open format that is NOT described in this proposal, which seems to be counter to the earlier principle that the government should not oblige any citizen or business to incur additional costs in order to share and collaborate with government documents. Accepting OpenXML into this list will enable the government to follow this mandate with the largest proportion of citizens and businesses (whether they use Microsoft Office or not), because it will include anyone who can deal with OpenXML, ODF and HTML - a cohort of the constituency which must be close to including everyone.

Documents may be shared in other formats but only in response to a specific request from a user

This should be an action reserved for only the least popular or esoteric of formats. OpenXML is neither of these. It is, in fact, the most popular of the modern open (XML-based and compressed) document formats.

Existing documents should be migrated to the formats specified in this proposal if they are re-opened for editing or are requested by a user

If the government follows our recommendation to include OpenXML in its list of standards, then either much of this work would be unnecessary or the cost of converting from older formats (e.g. from Microsoft Office Binary formats) would be greatly reduced or even completely eliminated.

Government bodies should avoid bespoke implementations which may limit their ability to migrate information or to share it with other users

Such a principle may seem attractive in general, but clearly begs the question "a bespoke implementation of what?" There may be good reasons for avoiding bespoke implementations of anything for which there is a commodity, off-the-shelf alternative ("why build, not buy?"), but sometimes a unique piece of business process logic may need to be written in order to address a unique business requirement (e.g. the UK's income tax or benefits system is unique to the UK, existing, as they do, in single systems operated by and on behalf of a single customer). So long as the implementation is proven to adhere to the specification of any standard chosen for use within it, there should be no migration problems for data written to that standard. For example, most government web sites are bespoke to one extent or another (they certainly appear nowhere else on the Internet without a degree of modification), but they are written to comply with standards (like HTML) which are chosen to ensure interoperability (largely, in the case of web sites, with as many web browsers as it is practical to support). If, for example, the tax system produced information to be consumed (viewed or edited) by a taxpayer (citizen), the system could be written to produce OpenXML (or ODF or PDF) directly by following the standard, without the need to produce the content in an intermediate format and convert it into an open standard format before sending it to the citizen. This is the benefit of open standard – there is no limitation on what sort of tool (application, app, program or service) can use them. Just because a standard describes a document format does not constrain the software that produces the document to being a word processor.

Macros should be avoided wherever possible, particularly when sharing documents.

Macros are incredibly useful, but also a subject of some debate. In broad terms, they are simply a means to an end and there are a number of circumstances where their use is either invisible to the user or benign to all the recipients of a particular document. Again, the capability to process a macro depends more on the capabilities of the tool (application, app, program, service) used than on the format chosen. If an author depends on the widest reception of their document - and use of the macro therein - then it is clear that with a range of tools available to users (not to mention devices and their inherent processing capabilities chosen by users), the author cannot guarantee that the macro can be processed as intended. However, to set a rule which says "macros should be avoided" (even with the "wherever possible" rider) risks inadvertently disabling their use in situations where all

collaborators are known to have access to a tool which will process the macro or where it is explicitly declared in the document that a macro needs to be run, guiding the recipient in their choice of tool (and, perhaps device), either of which many be the most effective and least cost means of getting the job done.

Government officials should engage with interoperability testing initiatives for document formats

We would wholeheartedly encourage this activity across all the leading open standard formats and specifically for OpenXML, ODF, PDF and HTML for one very good reason - that no standard will wholly survive the ingenuity, creativity or innovation of the users of any tools that implement the standard when they create content to share with anyone else. Interoperability testing is more effective when tools and standards are exposed to real world users and uses.

Government officials should engage with standards bodies associated with the maintenance of standards that are agreed for document formats for use in government

We would wholeheartedly encourage this activity too, but caution that the depth of expertise needed to effectively contribute to any number of these technical standards may not be widespread within any organisation, government being no exception. This is less a caution and more an encouragement to understand the necessity of making an appropriate investment in skills and time in order to add and extract value from these processes, which can last for several years.

This proposal, if agreed, would apply to information produced by or on behalf of central government departments, their agencies, non-departmental public bodies (NDPBs) and any other bodies for which they are responsible. These government bodies would need implementation advice to give clarity about when to use particular formats, the user needs they meet and the interoperability that can be expected.

We would observe that whatever choices are made by central government, the effect of those choices, particularly in this subject area, will be felt much wider than the organisations strictly within this scope. The impact of this decision will extend to all parties - citizens, businesses and other parts of the public sector - whether that is the government intention or otherwise. Such widespread impact should be considered explicitly.

A document metadata profile is outside the scope of this proposal, although this may be the subject of other challenges taken through the Standards Hub process.

Microsoft welcomes the opportunity to consider any future challenge on metadata.

Assessment of tools that can be used for providing multiple formats from a single standardised format are also outside the scope of this standards challenge.

We would again recommend that the government does not use a single standardised format, but embraces and supports a small range of formats which it may then choose from appropriate to the user need. When selecting tools that can used for conversion between

formats, the government should clearly assure itself of any tool's claimed compliance with a standard, but the government should also consider carefully the total cost of ownership for tools that would be widely used by officials. To cut to the chase, license costs alone are not a reliable indicator of the Total Cost of Ownership (of any IT system) and a thorough and robust TCO analysis should be as much a part of any procurement decision as an assessment of the functional capability of the tool itself against the functional requirements or user needs.

The remaining sections of our response cover "Facts about standards", "Data on the use of popular formats", "Application support for different formats", "The dynamic nature of standards", "Impact on the wider economy" and "Conclusions".

Facts about standards

ODF is an open international standard for document formats with at least two Standards Development Organisations (SDO). By OASIS as ODF v1.2 and by ISO/IEC as ISO/IEC 26300:2006 (also known as ODF v1.1).

ODF standards are governed by ISO/IEC Joint Technical Committee 1, Sub-Committee 34, Working Group 6 (or ISO/IEC JTC1/SC34/WG6)

ODF v1.2 is not yet an ISO/IEC standard, but will be considered by the ISO/IEC JTC1/SC34/WG6 in due course.

Open XML is an open international standard for document formats with at least two Standards Development Organisations (SDO). As ECMA-376 and as ISO/IEC 29500-1:2012.

ISO/IEC 29500 has four parts (ISO/IEC 29500-1, -2, -3, -4:2012) and allows for two states of implementation commonly known as Open XML "transitional" and Open XML "strict". These are not separate standards and are fully disclosed in the ISO/IEC 29500 documentation.

OpenXML standards are governed by ISO/IEC Joint Technical Committee 1, Sub-Committee 34, Working Group 4 (or ISO/IEC JTC1/SC34/WG4)

Microsoft Office 2007 supports ODF v1.1 and Open XML "transitional"
Microsoft Office 2010 supports ODF v1.1 and Open XML "transitional"
Microsoft Office 2013 (and Office 365) supports ODF v1.2 and Open XML "transitional" and Open XML "strict"

The current versions of LibreOffice and OpenOffice support ODF v1.2 and Open XML "strict".

Older versions of Open Office and LibreOffice support ODF and Open XML.

PDF v1.7 is a superset of PDF standards described collectively in ISO 32000-1:2008, which includes related standards like ISO 19005-1:2005 (or PDF/A, which is also PDF v1.4). PDF v2.0 is in development (by Adobe) and, as such, is not yet ready for consideration by ISO Technical Committee 171, Sub-Committee 2 (or ISO/TC 171/SC 2)

Microsoft Office support for PDF is explained in the following URL:
http://blogs.msdn.com/b/officeinteroperability/archive/2013/04/04/microsoft-support-for-pdf.aspx.

Microsoft has published all its old Binary document formats under its Open Specification Promise, making them available free of charge to anyone who chooses to still implement them. This means that while some may do, nobody actually needs to "reverse engineer" these formats, as their specifications are freely available.

Microsoft Office is the most popular office productivity suite both globally (with over one billion users) and in the UK.

Data on the use of popular formats

Searching on http://www.google.co.uk for the frequency of different document formats in different Internet domains revealed (on 5th February 2014) the following:

In all domains there are more .doc, .xls and .ppt files than any other editable format, which is unsurprising given that these formats have been in the market for over twenty years.

In the **gov.uk** domain there are 182 times as many DOCX files as ODT files, 12 times as many XLSX spreadsheets as ODS spreadsheets and 34 times as many PPTX presentations as ODP presentations.

In the **nhs.uk** domain there are 13,000 times as many DOCX files as ODT files, 1,910 times as many XLSX spreadsheets as ODS spreadsheets and 1,620 times as many PPTX presentations as ODP presentations.

In the **org.uk** domain there are 124 times as many DOCX files as ODT files, 61 times as many XLSX spreadsheets as ODS spreadsheets and 129 times as many PPTX presentations as ODP presentations.

And in the **co.uk** domain there are 133 times as many DOCX files as ODT files, 50 times as many XLSX spreadsheets as ODS spreadsheets and 108 times as many PPTX presentations as ODP presentations.

The table below shows the number of items found by Google's search engine in each domain for each file type. File types are grouped into those related to textual documents (PDF, DOC, DOCX, ODT, TXT, RTF), those related to spreadsheets (XLS, XLSX, ODS, CSV) and those related to presentation graphics uses (PPT, PPTX, ODP).

Domain	PDF	DOC	DOCX	ODT	TXT	RTF	XLS	XLSX	ODS	CSV	PPT	PPTX	ODP
gov.uk	12,400,000	2,420,000	43,400	238	41,800	127,000	561,000	9,640	780	37,900	29,300	1,690	50
nhs.uk	2,980,000	698,000	13,000	1	5,260	2,220	28,300	1,910	1	1,140	12,500	1,620	1
org.uk	14,500,000	4,130,000	261,000	2,100	3,180,000	109,000	191,000	18,700	307	35,700	163,000	16,300	126
co.uk	9,890,000	4,340,000	456,000	3,440	2,370,000	335,000	76,400	38,800	774	44,000	184,000	27,800	258

Similar results are produced by Microsoft Bing, but we have used Google to avoid any suggestion of a conflict of interest.

These results (dynamic and "live" as they are) illustrate (at this point in time) that the most popular current "revisable" format used by people across all four domains for each use type is OpenXML.

We would add the following notes:

1. Strictly speaking, PDF is also a "revisable" format, but it is most often used to create documents for viewing. No-cost viewers are available for almost every device and platform allowing a PDF document to be viewed (read) and in some cases even allowing the reader to bookmark or annotate their copy of the document with comments. In general, PDF readers do not allow the user any ability to modify the original content received.

2. According to the results above, Microsoft's old Binary formats actually remain the most popular "revisable" formats in use. Re-running the search above after even just a few hours showed that the number of items found in these old formats had increased. However, our focus here should be on "current" formats, i.e. those specified in an up-to-date, modern and international "open" standard. And while Microsoft has made these old Binary formats available at no cost under our Open Specification promise (one could describe them as "de facto open" formats), they are not the result of an open standardisation process and nor are they based on an extensible mark-up language (XML) schema in the way that both OpenXML and ODF are.

Even before Microsoft made its binary formats freely available under our Open Specification Promise (http://msdn.microsoft.com/en-us/library/gg615407(v=office.14).aspx), Microsoft's binary file formats had become (and seem to remain) the most widely used revisable document formats. Since publication in 2011, developers of other office productivity suites need not "reverse engineer" these formats. These formats still enjoy the widest range of use in both on-line services (some at no cost) and in applications which can be installed locally on a variety of devices (some at no cost). Such applications are all licensed in one way or another, whether use of the license is paid for or not and whether their source code is openly available or not, giving people tremendous choice in how the create and consume a variety of content.

Application support for different formats

There appears to be no authoritative source for the comparison of personal productivity applications and their capabilities, but perhaps, even as problematic as it might be, this reference on the ubiquitous Wikipedia (http://en.wikipedia.org/wiki/Comparison_of_office_suites) offers some indicative comparisons.

Following historical use of Microsoft's Binary formats, which appears to make them remain the de facto first choice for applications seeking the widest compatibility, it seems that the second most widely used format (or format family) across such applications is (arguably) OpenXML, with ODF a close third. Several of these applications are notable for their "genealogy", particularly amongst the open source products, where several applications now exist because of earlier "forks" in the code. Some of the branches have died out (e.g. KOffice), with others persist (e.g. forks of the now defunct Sun StarOffice exist now as Apache OpenOffice and LibreOffice). Some proprietary products listed on reference above have left the market (e.g. Microsoft Works, which Microsoft discontinued in late

2009 and Lotus SmartSuite, which IBM ceased to market in June 2013 and for which all support will cease in September 2014), while others continue (e.g. Corel WordPerfect Suite). While the contents of the reference are not perfect (or up-to-date), it does show some broad trends.

For most of these applications, there is a continuing and building theme to support both OpenXML and ODF. The most recent versions of OpenOffice and LibreOffice both claim additional development and interoperability testing for "round trip" OpenXML support.

It is also notable that some on-line services have recently chosen to focus only on OpenXML, abandoning ODF support (e.g. Google Docs). And we note also that a number of free-to-use and paid-for applications (or "apps" as they've become known in many tablet or phone on-line stores) that support a range of mobile device platforms do not support ODF and have chosen to implement OpenXML because of its inherent popularity.

This would indicate that the market is still highly active and that support for these formats (and the achievable interoperability across the implementations of any format) is not guaranteed in the long term. Things do change and it is difficult to predict either how or when.

The dynamic nature of standards

To illustrate this point, in the section "The dynamic nature of standards" below, we take a closer look below at a format which is widely used now, but not yet "standardised" (as in approved and maintained by an appropriate Standards Development Organisation or SDO), but which may become popular in the future. This is both the nature of the market and the process of establishing open standards. The lesson would seem to be that it is wiser to avoid limiting one's self to a narrow selection of standards.

It then follows that the safest way forward for any organisation seeking to ensure it had access to the maximum range of interoperability, the richest range of functionality and the widest common use of formats should be to seek to embrace multiple document formats.

Specifically, we would again recommend ODF (ISO/IEC 26300-1:2006 or OASIS ODF v1.1, with a path to accommodate the use of OASIS ODF v1.2 when ratified as an ISO/IEC standard), OpenXML (ECMA-376, ISO/IEC 29500-1:2012 "transitional" and "strict") and PDF (ISO 32000 and its associated standards like ISO 19005 or PDF/A).

In addition, we also concur with the proposal that many use cases can be best accommodated using a web-based HTML form, with W3C's HTML v5 specification being the most up-to-date version of that standard.

User adoption and application support then offers the prospect of the lowest operational cost in the long term and the greatest ability to provide or receive documents in a format that the citizen may prefer.

Contrary to the assertions of a number of commentators, both OpenXML and ODF remain vibrant and live standards in active development. Nowhere is this revealed better than in the latest

Business Plan from the ISO/IEC JTC1/SC34 ("SC34"), of which WG4 covers the development of ISO/IEC 29500 (OpenXML) and WG6 covers the development of ISO/IEC 26300 (ODF).

(The latest SC34 Business Plan can be found at http://isotc.iso.org/livelink/livelink?func=ll&objId=15737505&objAction=browse&viewType=1)

SC34's Business Plan for the Period October 2013 to September 2014 clearly details the close and cooperative working relationships between OASIS and WG6 for ODF and between ECMA and WG4 for OpenXML. Achievements listed for the prior period (up to September 2013) include (quoting from the plan):

- "*SC 34 has been actively maintaining ISO/IEC 29500. WG4 has started a project for revising ISO/IEC 29500-3 in reply to far-reaching defect reports, and will publish a working draft in the very near future.*"
- "*WG 6 experts have assisted the OASIS ODF Technical Committee in work to consolidate the technical alignment of ISO/IEC 26300:2006 with Amendment 1 and ODF v1.1. Over the past year some further minor defects have been is covered in ODF v1.1, OASIS has published corrections to their own specification and work has now started on correcting ISO/IEC 26300 to align with these corrections.*"

Both statements show that ODF and OpenXML are being actively maintained and improved by these communities.

Indeed, the Business Plan also notes that "*SC34 looks forward to the anticipated PAS submission of ODF 1.2 during the next period.*", indicating that while SC34 fully expects ODF v1.2 to update the ISO/IEC 26300 standard, it does not expect ODF v1.2 to begin its journey through the ISO/IEC acceptance process until October 2014 at the earliest (i.e. the start of "*the next period*").

New standards are always being developed and sometimes these appear to offer similar functionality to existing standards, which themselves may also continue to be maintained and developed. A number of comments in the SC34 Business Plan relate to the standardisation of the specification of EPUB (in collaboration with the International Digital Publishing Forum or IDPF). EPUB is XML-based digital publishing standard for e-books.   If nothing else, consideration of this standard demonstrates that the ecosystem and enthusiasm for new document formats is healthy and vibrant, with this standard perhaps becoming a potential candidate for "viewing government documents" at some future time - a possibility that would place EPUB seemingly in direct competition to PDF.

Yet this illustrates an important conclusion we seek to emphasise here - that international standards can and do happily co-exist even where they seemingly compete. While those that do appear to compete cover many of the same "use cases" for similar groups of users, each also often exhibit unique features or user benefits that are both valuable and valid to retain.

At some future time the UK government may ask people to consider (or, for that matter, the people may ask the government to consider) the adoption of the EPUB standard for viewing government documents. At that time it will have to ask whether to do so would invalidate or deprecate the need for PDF. The answer should be a resounding "no". EPUB and PDF may offer greater benefits

together to users (citizens and civil servants) for no significant or measurable increase in operating costs for the government. Applications that the government might then choose may (or should) be able to cope with both formats covering each format's shared and unique features.

The same is true right now and in a completely analogous way for document formats for sharing and collaboration. ODF and OpenXML both have their place and their preferred usage - and that is a preference both of users (having chosen that format) and uses (being capable in some different and valuable ways).

Impact on the wider economy

We believe that including OpenXML, along with ODF and HTML, in the standards for sharing and collaborating with government documents will allow the government to achieve all the objectives it has set out when setting this Challenge. Indeed, we would say that it is more able to deliver what users need only if OpenXML is included, because it widens choice and interoperability opportunities across all IT systems, not merely the choice of a small number of specific productivity applications. Consideration of this latter outcome demands that before a final decision is taken, the government should appraise (and share that appraisal) the impact on wider IT economy of making the current proposed narrow selection of ODF only.

We support and applaud the government's G-Cloud programme and the progress it has made to opening up government IT business to small and medium sized enterprises (SMEs). Microsoft's Partner ecosystem comprises over 35,000 partner organisation in the UK (and over half a million partners worldwide), employing over 250,000 people in the UK alone. Our partners tell us they are encouraged and excited by the developments the government is making its markets more open and accessible to them. The overwhelming majority of these partners are SME.

A significant proportion of these partners build or develop applications that run on or work with our products and services. The role of our support for open standards across our products, platforms and services is a factor that allows our partners to create interoperable products and services they can then sell to their customers across the UK and further afield – including into the Public Sector.

While Microsoft Office supports a wide range of open standards, including OpenXML and ODF, our partners have a choice of which open standards they use in their products.

By making such a constrained choice, as in the current proposal, the government risks either alienating many UK SME businesses who have based their products, quite freely and reasonably, on open standards like OpenXML.

OpenXML is not merely a standard for word processors, with documents only ever created and read in one or other word processor, it can be used by applications to assemble and communicate complex information which will eventually be read by people, but not necessarily by those people using a word processor. This is a decision that has impacts much wider than how many different word processors the government has to choose from.

We are surprised that throughout this consultation no mention or assessment has been made (other than by Microsoft) of the financial impact of choosing a narrow range of document format

standards. We believe such an assessment, taking into account the wider impact on the UK IT ecosystem – particularly the SME IT suppliers in the UK (who may not be able to afford the investment needed to change their product to make it attractive to government) – should be thoroughly considered, through an open process and made public before any final choice is made.

Conclusions

We have shown here that OpenXML enjoys a popularity of use across the major domains of interest to the government, including across the major domains of the UK Public Sector (gov.uk and nhs.uk) and of UK business (co.uk) and of the 3rd sector (.org.uk).

We have also shown that both standards (OpenXML and ODF) are vibrant and developing, that they are both supported by a range of tools (applications, apps, programs or services) widely available to the government, to business and to citizens in a range of circumstances to suite user needs.

And we have also shown that it is the nature of standards to develop and occasionally supersede one another, but that for a time, multiple standards that appear to do the same thing can and do exist and thrive in parallel for many years.

On this basis, we urge the government to avoid a costly and unnecessary focus on too narrow a selection of standards. Having already recognised in its proposal that no single standard will be adequate for its needs, we now urge the government to include OpenXML alongside ODF and HTML in its list of standards for sharing and collaborating with government documents.