TECHNICAL REPORT

ISO/IEC TR 22981

First edition
2020-06

Information technology — Office equipment — Guidelines for the development of an ontology (vocabulary, components and relationships) for office equipment

Technologies de l'information — Équipement de bureau — Lignes directrices pour l'élaboration d'une ontologie (vocabulaire, composantes et relations) pour l'équipement de bureau

Reference number
ISO/IEC TR 22981:2020(E)

Contents

Page

Foreword iv

Introduction v

1 Scope 1

2 Normative references 1

3 Terms and definitions 1

4 Key concepts and current activities 1

4.1 General 1

4.2 Virtualisation of physical objects and processes 2

4.3 Big data and artificial intelligence 2

4.4 Peer-to-peer communication and smart contracts 2

4.5 Identification and authorisation 3

4.6 Ontology development 3

5 Steps in the development of an [office] ontology 4

5.1 Selection of appropriate tools 4

5.2 Identification of elements 4

5.3 Relationship to existing ontologies 4

5.4 Other vocabularies and standards development organisations 5

5.5 Example of automated communication 6

5.6 Vocabulary specification 7

Annex A (informative) **Basic RDF concept** 8

Bibliography 10

Foreword

ISO (the International Organization for Standardization) and IEC (the International Electrotechnical Commission) form the specialized system for worldwide standardization. National bodies that are members of ISO or IEC participate in the development of International Standards through technical committees established by the respective organization to deal with particular fields of technical activity. ISO and IEC technical committees collaborate in fields of mutual interest. Other international organizations, governmental and non-governmental, in liaison with ISO and IEC, also take part in the work.

The procedures used to develop this document and those intended for its further maintenance are described in the ISO/IEC Directives, Part 1. In particular, the different approval criteria needed for the different types of document should be noted. This document was drafted in accordance with the editorial rules of the ISO/IEC Directives, Part 2 (see www.iso.org/directives).

Attention is drawn to the possibility that some of the elements of this document may be the subject of patent rights. ISO and IEC shall not be held responsible for identifying any or all such patent rights. Details of any patent rights identified during the development of the document will be in the Introduction and/or on the ISO list of patent declarations received (see www.iso.org/patents) or the IEC list of patent declarations received (see http://patents.iec.ch).

Any trade name used in this document is information given for the convenience of users and does not constitute an endorsement.

For an explanation of the voluntary nature of standards, the meaning of ISO specific terms and expressions related to conformity assessment, as well as information about ISO's adherence to the World Trade Organization (WTO) principles in the Technical Barriers to Trade (TBT), see www.iso.org/iso/foreword.html.

This document was prepared by Joint Technical Committee ISO/IEC JTC 1, *Information technology*, Subcommittee SC 28, *Office equipment*.

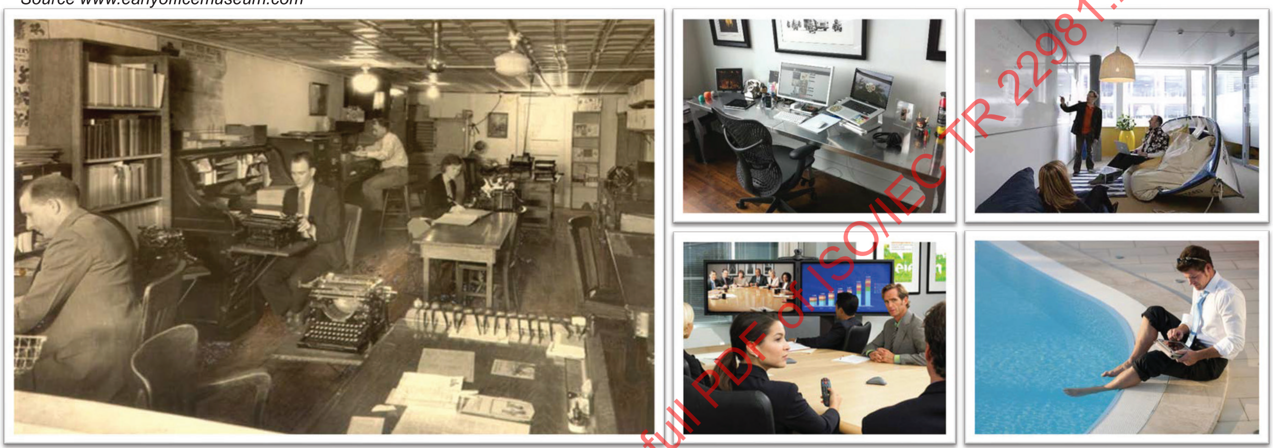
Any feedback or questions on this document should be directed to the user's national standards body. A complete listing of these bodies can be found at www.iso.org/members.html.

Introduction

The aim of this document is to provide an overview of the current status of 'office equipment', to identify likely future directions and to show how standardisation of terminology in the form of an ontology could provide a strong framework for office automation.

Historically 'office equipment' was more clearly defined than it is today. The concept of what constitutes 'the office' is changing and people no longer have to go to a particular place to perform tasks traditionally associated with the office. These tasks include activities such as creating or copying documents, reading reference books and preparing presentation materials can now be carried out in many different environments as shown in Figure 1.

Source www.earlyofficemuseum.com



Traditional office ***Some modern 'offices'***

Figure 1 — Ideas of what constitutes an office are changing rapidly

The set of tools and equipment traditionally available to office users has also expanded and now includes on-line document creation and sharing, high quality colour displays, digital projection equipment, video communication and so on. Indeed there seems to be a new class of equipment, used to enable tasks traditionally associated with 'the office', introduced each year.

Consequently, there is no longer a clear boundary between office, personal, industrial and commercial documents as many of the same design tools and reproduction equipment are used across all of these sectors. For example, Microsoft Publisher, Adobe InDesign and Microsoft Word are used to create company brochures, personal documents and product handbooks.

With the advent of high-speed internet, there is no longer a need for 'office workers' to be in the same location and, in many cases, companies operate a virtual office with members working from many different locations. To support this, documents are made available in a central repository, for example using a document management system, and can be viewed, modified and printed at any location as long as the user has the required rights.

When considering requirements for future office standards it is important to consider these recent developments and also to think about what users will expect to be able to do in future. Given the fact that the same range of systems and supporting software are used for several different purposes, the scope of these standards should also be considered. One option would be to focus on the set of tools used to communicate business ideas, particularly the creation, display, printing, copying, distribution and archival of documents.

In a Scientific American article in May 2001[4], Tim Berners-Lee proposed a concept that he referred to as 'the Semantic Web'. He envisaged the development of the internet in such a way that information could be identified by structured metadata, for example its type, creator, intended use and so on. In this article the authors observe that three of the important technologies are already in place: eXtensible Markup Language (XML)[26], the Resource Description Framework (RDF)[5] and Uniform

Resource Indicators (URIs)[27]. In its simplest form, XML allows the creation of tags such as <name> and <country> and associated values such as <John Smith> and <China> and these data names and values can be used to annotate web pages. While in some cases these can be understood by a person reading these pages, the meaning cannot generally be understood by a machine. The meaning, however, can be expressed using RDF which provides a mechanism to allow properties of particular things to be described in a way whereby they can be parsed unambiguously by machines. For example <name> <is a resident of> <country>, has a structure rather like the subject, verb and object of an elementary sentence. The subject and object can be defined precisely using URIs and with care URIs can also be used to define the relationship (the verb) between the subject and object. The authors observed that this model can be used to allow web data to be interpreted by machines. This is a very powerful concept and with only minor modifications is one of the basic principles for the Internet (or web) of Things (IoT).

The basic concept of the IoT is that all internet objects have a unique identifier such as an IP address, URL, URI, barcode or RFID tag. ‘Things’ include any kind of uniquely identifiable object including both physical and abstract objects such as a device, web page, communications link or an item with an RFID tag or barcode. Where suitable infrastructure and authentication exists, devices can communicate directly with other devices, for example a home security system can detect an intruder and pass this information to a mobile telephone to alert the home owner (device-to-device). Similarly, devices can obtain data from web pages, for example a mobile phone can find the weather forecast in Tokyo or a list of books published by Springer on colour management. Thinking about the needs for accurate colour reproduction, a digital printer can find a suitable ICC profile for its current configuration (paper, ink set, resolution etc.). This document uses identification of a suitable ICC profile as a practical example.

Ensuring that equipment used in the office of the future can communicate effectively with minimal user intervention is one of the goals of manufacturers of all such equipment. An important prerequisite to this kind of communication is the ability to be able to identify internet resources clearly using some form of standardised vocabulary or ontology[10].

One of the mechanisms proposed for the development of ontologies is RDF, the latest version of which was published by W3C on 25th February 2014[5]. Details of RDF and the way in which it is supported is complex and Annex A provides an overview of the principles. Further research is needed, particularly to investigate details of registration of standard predicates.

One benefit of RDF-enabled documents is that search engines can provide more useful search results as they can discover web resources of particular types and order these in a way that is useful to someone who wishes to browse the set.

In addition to the basic RDF syntax there is a need to standardise the types of things and the relationships between them. The W3C recommends the use of Web Ontology Language (OWL)[27] which provides the framework to define basic classes, properties, individuals, and data values which can be used in conjunction with RDF for a given domain. More recently (July 2017) W3C introduced the Shapes Constraints Language (SHACL) which provides constraints necessary for data validation. OWL is designed for classification tasks (inferencing in an ‘open world’), while SHACL covers data validation (in a ‘closed world’) in a similar way to that of traditional schema languages.

Perhaps it is easiest to understand how an ontology can be defined by looking at an example. Here, the ontology of BBC programmes[17] has been selected and a very high-level overview of this is shown in Figure 2.

The first observation is that this ontology is built on a number of others:

— The Music Ontology[18] provides main concepts and properties for describing music (i.e. artists, albums and tracks) in terms of, for example a 'Composition','MusicArtist', 'AudioFile' and so on. This ontology was developed independently and is widely used by other ontologies.

— FOAF (an acronym of Friend Of A Friend)[19] is a machine-readable ontology describing people, their activities and their relations to other people and objects.

— SKOS (Simple Knowledge Organisation System) is used to define some of the more abstract concepts such as topics of a program.

— Event ontology describes where and when the event was held and who the participants were.

These ontologies in turn build on other ontologies. The BBC ontology is part of a large web of connected ontologies which can be explored from the Linked Open Vocabularies website[15]. The BBC Programmes ontology defines specific classes such as 'Broadcast', 'FM', 'Format', 'Genre' and so on.

Image from http://www.bbc.co.uk/blogs/radiolabs/2008/03/whats_that_star_on_the_wall.shtml

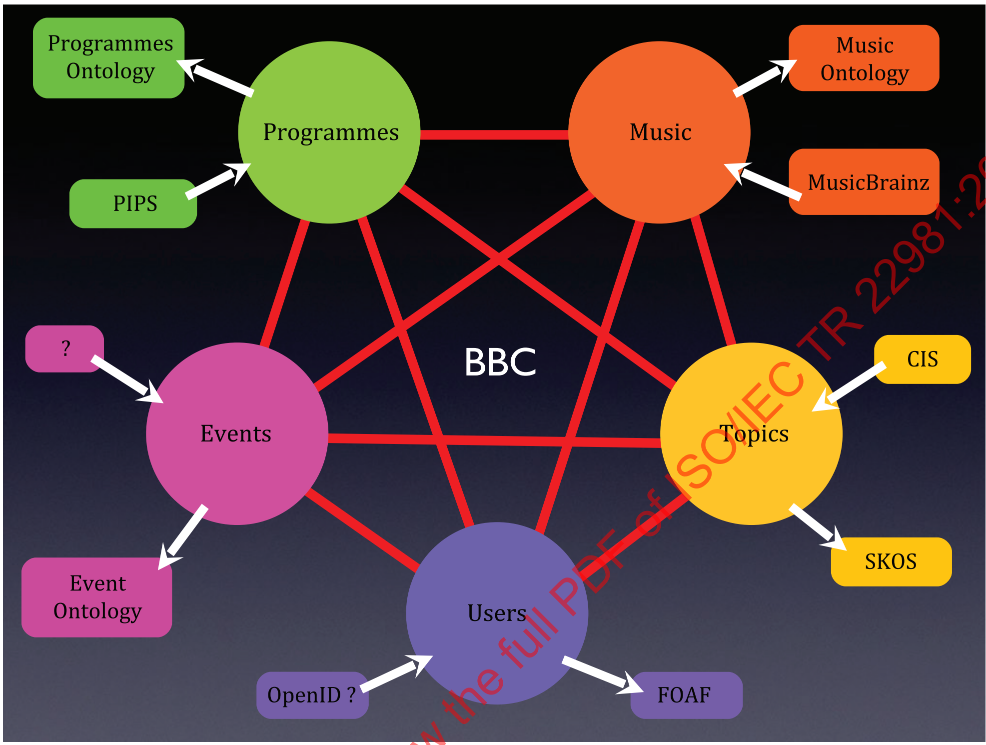


Figure 2 — An example of an ontology

In a similar way, it would be possible to build an ontology for office equipment, office documents, internet resources and so on. In order to provide a concrete example, the task of automatically identifying ICC profiles has been selected. This example addresses only one aspect of office devices (that of achieving colour fidelity) but is sufficiently complex to show how standardised metadata can be used in practice.

For all of these solutions to be useful, some kind of domain-specific ontology is needed. An example of the identification of ICC profiles is used in this document but this should be seen as part of a broader ontology, for example an imaging and print ontology or an office actors, tools and resources ontology. There are aspects of ICC profile metadata that are common to other areas, for example identification of colour imaging devices, print substrates, printing inks and other more abstract metadata such as identification of the copyright holder, creator and owner. This is an important consideration and where such metadata is already well defined, this should be incorporated by reference and not redefined by any new ontology. There are a number of groups closely related to that of office equipment and any work done should be coordinated with these other groups.

Information technology — Office equipment — Guidelines for the development of an ontology (vocabulary, components and relationships) for office equipment

1 Scope

This document provides background information and guidelines for the development of an ontology for office equipment.

An example of how such a standard can be used to automate the identification of resources for colour is provided.

NOTE Often the terms vocabulary and ontology are used with the same meaning.

2 Normative references

There are no normative references in this document.

3 Terms and definitions

For the purposes of this document, the following terms and definitions apply.

ISO and IEC maintain terminological databases for use in standardization at the following addresses:

— ISO Online browsing platform: available at https://www.iso.org/obp

— IEC Electropedia: available at http://www.electropedia.org/

3.1
office actor
person or machine conducting a specific role related to an office function

3.2
ontology
specification of concrete or abstract things, and the relationships among them, in a prescribed domain of knowledge

[SOURCE: ISO/IEC TR 19763-9:2015,3.1.3 modified — Note 1 to entry has been deleted.]

4 Key concepts and current activities

4.1 General

There are several important concepts that are emerging and seem likely to become important for future business technologies. Together these enable direct peer-to-peer communication and form the basis of emerging concepts such as smart cities, smart homes, smart cars, smart office and many others. These future models assume that many devices will be used, and these devices will communicate with each other directly rather than requiring a central control as is currently required. Such a central control becomes impractical when the number of devices increases to the extent envisaged.

Key concepts include virtualisation of physical objects and processes, big data, artificial intelligence, peer-to-peer communication, digital rights management, identification and authentication. The main topic of this document is the role played by an ontology but some context for this may be helpful.

4.2 Virtualisation of physical objects and processes

Virtualisation is not a new idea. Readers will no doubt be familiar with the representation of a printer by an icon on our computer's virtual desktop. We can control the behaviour of the printer by changing settings using a print dialog box and request that a virtual document created by interacting with our computer keyboard and other controls should be printed. This results in the creation of a physical copy of something that up to that point only existed as a virtual entity. Over recent years, this basic model has been expanded to provide virtual information about physical aspects of the printer such as the amount of ink remaining, the size and sometimes the type of media present and various other controls that adjust physical aspects of the printer in order to achieve the desired print appearance.

This basic concept of virtualisation has been expanded by the Industry 4.0 initiative originated by the German Federal Ministry for Education and Research[20] (popularised by Siemens and Bosch). In Industry 4.0, many aspects of industrial machines and processes have a virtual representation, and this enables intelligent machines to communicate directly with one another. For example, a robot creating a product as part of an assembly line can indicate that it is ready to begin a new assembly, this indication can be recognised by another device (an inventory controller) which can request further robotic systems to deliver the set of parts needed. Similarly, the inventory controller can detect when the supply of parts is low and automatically order new parts to be delivered.

The flow of processes on a production line can be monitored and optimised or modified based on changes in demand for the product being manufactured. In this way manufacturing operations can be fully automated, requiring intervention only when something goes wrong.

One key component of this virtualisation is the increased availability of smart sensors. Almost every aspect of an environment can be monitored such as energy usage, temperature and room lighting. Smart actuators are also becoming available enabling remote control of temperature, lighting and other devices. In this way, sensors and actuators are used in industry to allow operators to interact with a virtual model of physical production systems and indeed for production systems to interact with one another with minimal need for operator intervention.

An important prerequisite enabling machines to communicate with each other is a common language. This is one of the reasons for the increased interest in the development of ontologies for industrial devices and processes.

4.3 Big data and artificial intelligence

Virtualisation of systems leads to huge increases in the amount of data available about almost any system. Making this data available 'in the cloud' opens the possibility that trends can be spotted early and systems controlled much more effectively. For example, multiple failures of a particular component can be detected early and replacements organised thus reducing system downtime or elimination of collateral damage caused by these failures. Processing and managing such large amounts of data requires the development of new methods. Companies such as Google, Microsoft, Apple, Amazon, Facebook as well as open source initiatives such as Apache Spark and many others have developed platforms for analysing large amounts of data in order to make effective decisions.

Increases in the power of computer systems, combined with improved access to and communication of data has led to complex computer algorithms that are able to solve complex problems that are beyond the capability of conventional computer algorithms. There are many different approaches taken to solve these problems, but all are included in the rather ill-defined term 'artificial intelligence'.

This combination of data science and artificial intelligence leads towards the goal of machines making high-level decisions based on simple directions from human controllers. It is obviously important that the available data can be interpreted by machines and to do so a common language is needed.

4.4 Peer-to-peer communication and smart contracts

One of the difficulties faced when many devices need to communicate effectively with other devices on a peer-to-peer basis is that of ensuring the integrity of such a network. It is impractical for a central controller to supervise the interactions between devices and a model that allows devices to communicate and interact directly with each other is required. This model should ensure secure communication between devices that are authorised to communicate and allow for a pre-determined range of interactions. One such mechanism is a new concept using distributed ledger technology which obviates the need for a single central controller. Perhaps the most well-known of these is the blockchain.

In 2008 Satoshi Nakamoto (an unknown person or group of people) developed the concept of the blockchain[21]. One of the first commercial products based on this technology was the Bitcoin cryptocurrency and this was followed by a number of other cryptocurrencies that use variations of the basic method described by Nakamoto. The blockchain concept, is not limited in application to cryptocurrencies but in principle could form the basis for peer-to-peer smart contracts and could become the primary mechanism for authentication of communication between machines.

This is a relatively new technology that could have huge potential, however further development is likely to be necessary before this technology becomes mainstream. Blockchain and similar technologies are the subject of a very large amount of research, one interesting development is that of Ethereum which reduces the total energy required to operate the system[22].

A common language is needed to support this level of direct peer-to-peer communication.

4.5 Identification and authorisation

Devices and individuals need a standard mechanism to allow other devices and individuals to know that they are who they say they are. Many new technologies are emerging in this area including the use of biometrics that use fingerprints, faces or walking rhythm to identify people. Similar mechanisms are needed for identification of machines in such a way as will allow machine-to-machine identification. These aspects are addressed in part using Subscriber Identification Modules (SIMs) and Machine Identification Modules (MIMs).

A mechanism is necessary to allow access to specific resources owned by a device or person. One scheme that is becoming widely used for this purpose is Open Authentication (OAuth) where trusted parties can share authentication of individuals.

As with the previous technologies, a common language is needed.

4.6 Ontology development

When these and other concepts are considered, it is clear that the development of an ontology is necessary, however there is no universal agreement as to how this should be developed. This section outlines some important principles and current activities.

Many of the standards in this area are developed by the World Wide Web Consortium (W3C) and the W3C data activity group[23] is responsible for the development of standards to facilitate potentially Web-scale data integration and processing. This group has published a document describing best practices for data on the Web[24].

The work of the Schema.org W3C community group should be considered as a possible model to follow.

A large set of more than 600 Linked Open Vocabularies[15] (LOV) is published on-line with easy-to-use search tools. This provides a useful resource in identifying existing terms in order to avoid their duplication.

To make the web of data a reality, it is important to have the huge amount of data on the Web available in a standard format, reachable and manageable by Semantic Web tools. Furthermore, not only does the Semantic Web need access to data, but relationships among data should be made available, too, to create a web of data (as opposed to a sheer collection of datasets). This collection of interrelated datasets on the Web can also be referred to as linked data.

Two important W3C Recommendations for the development of ontologies are RDF[5] and JSON-LD[25].

5 Steps in the development of an [office] ontology

5.1 Selection of appropriate tools

Before any work is done in this area it is important to agree how the data will be held and managed. The main reasons to do this is that it can be very difficult to move data from one development environment to another and it is likely (hoped) that others will link to the data defined by the office equipment community.

One popular tool is Protégé from Stanford University but there are numerous other options. Figure 3 shows two examples of the development and deployment process.

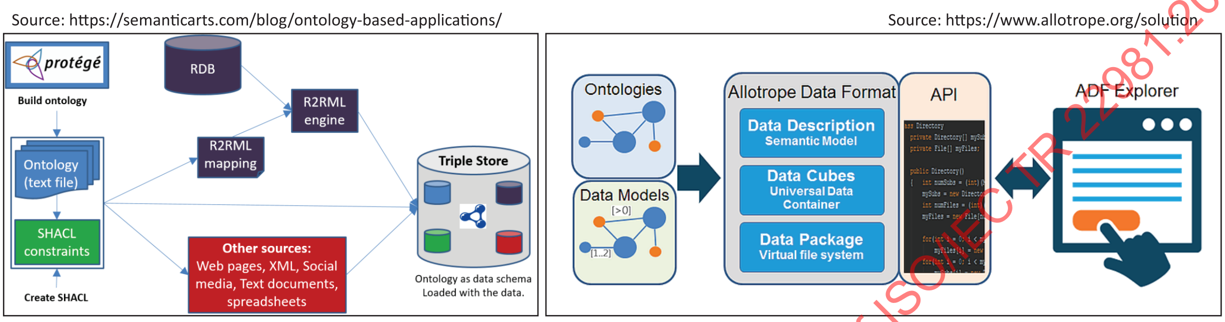


Figure 3 — Ontology development and deployment

The two diagrams in Figure 3 show two quite different development and deployment processes. The details of these processes will not be discussed here but note how the definition of ontologies fits with these two processes. An ontology can be considered a foundation technology that underpins a broad range of development processes.

5.2 Identification of elements

The first step is to identify the key elements of an ontology and explain each briefly providing examples where appropriate. At this stage, no classification is performed although it may be helpful to show some of the important relationships between elements and it may be helpful to start to group them in some way.

The following is a very basic start towards a list of the kinds of elements that should be considered:

Printers, cameras, displays, projectors, print substrate, ink, communications interface, ICC profile, resolution, printing technology, halftone method, connected, transmitter, receiver, calibration, colour channels, coating, status, dimension, copyright holder, owner, creator, creation date and many more.

5.3 Relationship to existing ontologies

As the list of elements is constructed it is important to identify any existing ontologies that already define these elements and to determine whether that ontology should be used or whether the intended usage is different from that of office systems.

Table 1 shows a few ontologies which may be relevant. Readers are strongly advised to review the work of the LOV community[15]. This community has developed a framework where ontologies and relationships between them can be identified and where users can search for metadata in order to determine whether a suitable definition already exists before creating a new definition.

Table 1 — Base ontologies that may be relevant when developing an ontology for office equipment

Name of owner	Description	Home page	Examples of what the ontology defines
Dublin Core Metadata Element Set (dce)	The Dublin Core Metadata Element Set is a vocabulary of fifteen properties for use in resource description. The name "Dublin" is due to its origin at a 1995 invitational workshop in Dublin, Ohio; "core" because its elements are broad and generic, usable for describing a wide range of resources.	http://dublincore.org/documents/dces	Agent, AgentClass, FileFormat, Frequency, Location, MediaType, PeriodOfTime, PhysicalMedium, PhysicalResource, RightsStatement, SizeOrDuration
Friend of a Friend (foaf)	FOAF is a project devoted to linking people and information using the Web. Regardless of whether information is in people's heads, in physical or digital documents, or in the form of factual data, it can be linked	http://www.foaf-project.org/	Agent, Document, Group, Image, Organization, Person
Time Ontology (time)	This vocabulary defines temporal entities such as time intervals, their properties and relationships.	http://www.w3.org/TR/owl-time	Interval, DateTimeDescription, Duration
For full details of these and many other ontologies, see https://lov.linkeddata.es/dataset/lov/.			

5.4 Other vocabularies and standards development organisations

Some ontologies that are important for office equipment can be applied to multiple domains whereas others cannot. It will be important to determine which aspects of an office ontology are likely to overlap with other domains and in these cases, it will be very important to work with joint partner organisations.

Table 2 shows existing industry vocabularies that are sufficiently well defined and have broad adoption within the office systems industry sector or in related sectors.

Table 2 — Existing industry vocabularies that should be considered when developing an ontology for office equipment

Name	Description	Home page	Defines (examples)
Job Definition Format (JDF)	Language for describing a print job across enterprises, departments and software and systems.	https://www.cip4.org/	Many standard print-related terms
eXtensible Metadata Platform (XMP)	Used to communicate metadata for digital documents and data sets. May be embedded in TIFF and PDF documents.	https://www.adobe.com/products/xmp.html	Digital document components and properties
Metadata Working Group, EXIF, IIM (old IPTC) metadata	Used to communicate metadata for images		Digital image properties
Printer Working Group (PWG)	The PWG is chartered to make printers, multi-function devices, and the applications and operating systems supporting them work together better.	https://www.pwg.org/namespaces.html	Printer, PrinterStatus, PrintJob, ColorSupported, PagesPerMinute, PrinterInfo, PrinterLocation, and many more

Table 2 *(continued)*

Name	Description	Home page	Defines (examples)
Printer Vocabulary Language Reference	Vocabulary to describe home and office printers including all-in-one devices with faxing, scanning and copying capabilities.	http://www.ebusiness-unibw.org/ontologies/opdm/printer.html	AllInOneInkjetPrinter, AllInOneLaserPrinter, Color, ColorRange, CommInterface, CopyMachine,, Fax, InkCartridge, InkjetPrinter, LaserPrinter, MatrixPrinter, Printer, Scanner, ScannerCopierInkjetPrinter, ScannerCopierLaserPrinter, Toner

Figure 4 shows some of the organisations that have developed terminology that overlaps in some way with that of the office activity.

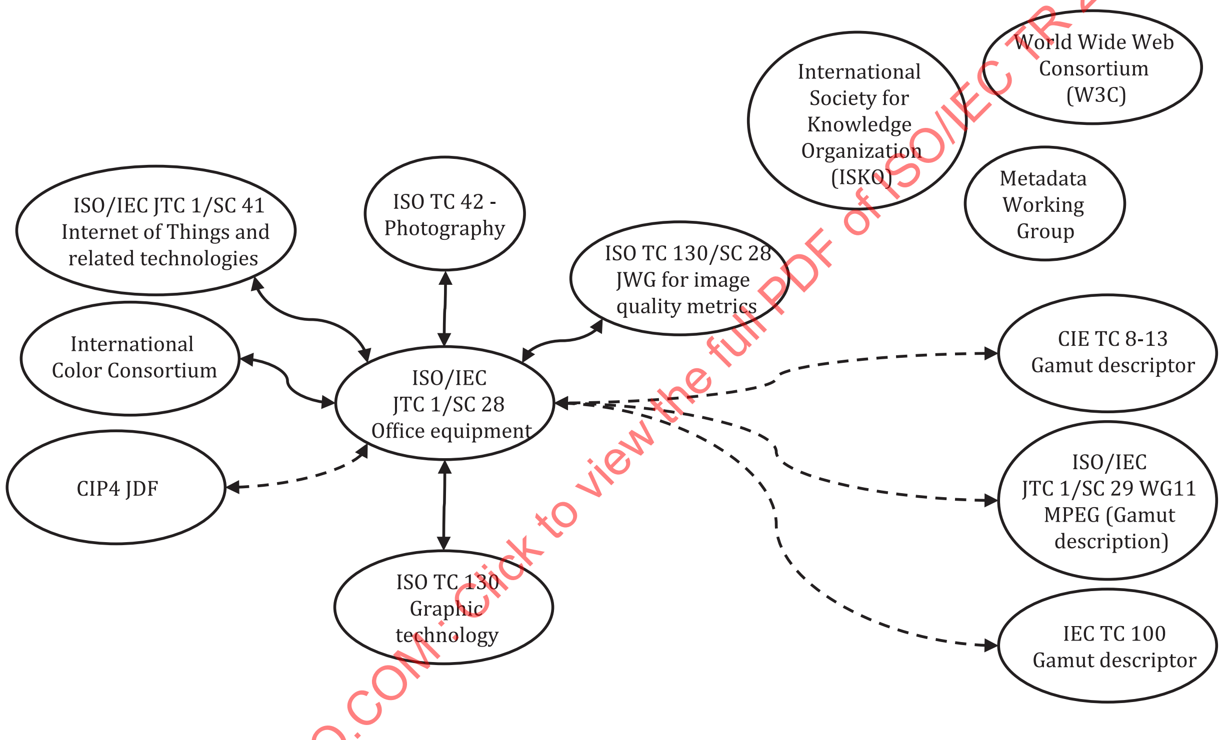


Figure 4 — Map of organisations which have developed terminology that overlaps with that of office equipment

The development of an ontology should be done in close collaboration with all of the groups shown in Figure 4. In this figure solid lines show groups with a close connection to office equipment, dotted lines show groups where there is some overlap and the groups with no lines currently have no connection with office equipment.

5.5 Example of automated communication

Figure 5 shows an example of a possible model for home and office printer connection and an effective mechanism to achieve high quality colour management. The figure shows the supplier of an ICC profile for a printer configuration uploading a set of ICC profiles (Device Profile Supplier). These profiles may be for a large set of printers and for different printer configurations (inks, media and printer settings). The figure also shows the equipment manufacturer who may qualify each profile. The office and home users wish to be able to select the most appropriate profile for their printer configuration from the repository.

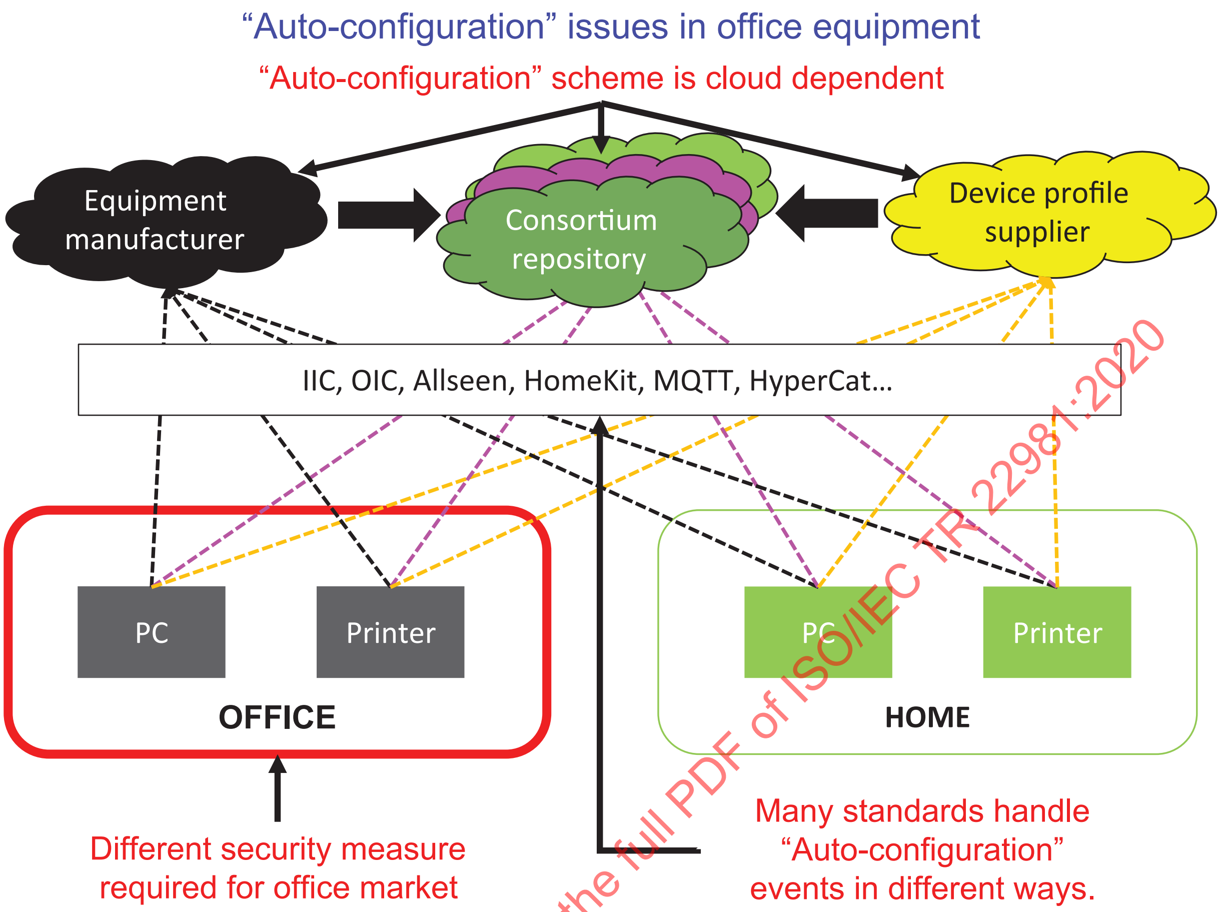


Figure 5 — Auto-configuration issues for office equipment

This figure shows a number of problems that should be addressed. There are many auto-configuration standards that all work in different ways and some common language is needed to allow systems like these to communicate with each other. The way in which security aspects are handled differs widely which may mean the ICC profiles cannot be downloaded effectively.

These communications can be improved by the development and adoption of a common vocabulary and ontology.

5.6 Vocabulary specification

The final step is to define elements specific to office systems. A key part of the development and deployment of an ontology is that it should be easily available to those who wish to implement systems based on its content. It should also be easy to revise when updates or corrections are needed.

Annex A
(informative)

Basic RDF concept

A.1 General

This annex provides a very brief introduction to the RDF concept. Readers are encouraged to refer to on-line tutorials such as those from w3schools[28].

A.2 RDF graph

An RDF graph is a section of XML that expresses a relationship between a subject and an object. It comprises a subject node, a predicate and an object node.

The subject and object nodes may be one of three types:

a) An IRI (Internationalized Resource Identifier): which is a more general form of a URI. This could be a web page, a device, a file and so on.

b) A literal: values such as strings, numbers, and dates.

c) Blank: which may be used to indicate that a relationship exists without naming the subject / object.

The predicate is an IRI and denotes a relationship or property, for example 'is called', 'is owned by', 'was created on', 'lives with', 'for use with' and so on.

A.3 Example

In this highly simplified example we wish to post an ICC profile (icc_profile1) in a way that can be identified clearly. In this case the ICC profile is for a printer 'printer_model1' and uses inks of type 'ink_set1'. Clearly more parameters are needed to identify the printing condition but for now let us assume that the only optional parameter is the ink.

The subject is:

— An ICC profile that can be identified by a URI, for example http://www.color.org/profiles/icc_profile1.icc.

The objects are:

— The printer model (with a predicate "for use with printer model"), this may be any IRI that identifies the printer model unambiguously, for this example we will use: www.printerwebsite.com/printer_model1.

— The printing inks (with a predicate "for use with printing inks"), in this example this is an IRI that identifies the ink set: www.inkwebsite.com/ink_set1.

The predicates should also have URIs. For example, the URI for each predicate:

— for use with printer model could be http://www.color.org/rdf/usewith#printer,

— for use with printing inks could be http://www.color.org/rdf/usewith#inks,

So the example can be expressed in RDF/XML format as: