# ISO/IEC 27038

Search

● Search this site

Home | ISO27k standards | FREE ISO27k Forum | FREE ISO27k Toolkit | FREE ISO27k FAQ | DONATE!

ISO/IEC 27000
ISO/IEC 27001
ISO/IEC 27002
ISO/IEC 27003
ISO/IEC 27004
ISO/IEC 27005
ISO/IEC 27006
ISO/IEC 27007
ISO/IEC TS 27008
ISO/IEC 27010
ISO/IEC 27011
ISO/IEC 27013
ISO/IEC 27014
ISO/IEC TR 27016
ISO/IEC 27017
ISO/IEC 27018
ISO/IEC 27019
ISO/IEC 27021
ISO/IEC TS 27022
ISO/IEC TR 27024
ISO/IEC TS 27028
ISO/IEC TR 27029
ISO/IEC 27031
ISO/IEC 27032
ISO/IEC 27033
ISO/IEC 27034
ISO/IEC 27035

< Previous standard ^ Up a level ^ Next standard >

## ISO/IEC 27038:2014 — Information technology — Security techniques — **Specification for digital redaction** *(first edition)*

### Abstract

> *"ISO/IEC 27038:2014 specifies characteristics of techniques for performing digital redaction on digital documents. It also specifies requirements for software redaction tools and methods of testing that digital redaction has been securely completed. ISO/IEC 27038:2014 does not include the redaction of information from databases."*
>
> *[Source: ISO/IEC 27038:2014]*

### Introduction

Digital data sometimes have to be revealed to third parties, occasionally even published to the general public, for reasons such as disclosure of official documents under Freedom of Information laws or as evidence in commercial disputes or legal cases. However, where it is deemed inappropriate to disclose certain sensitive data within the files (such as the names or locations of people or sources who must remain anonymous and various other personal or proprietary information that must remain strictly confidential), those must be securely removed from the files prior to their release. 'Redaction' is the conventional term for the process of denying file recipients knowledge of certain sensitive data within the original files.

Given that redaction is usually relevant to the protection of highly confidential information, failures in the process that lead to inappropriate data disclosure are almost bound to be serious and in the worst cases can be grave. Redaction failures have led to incidents such as identity theft, disclosure of confidential security matters, privacy breaches and compromising the identities of undercover agents and informants, while disclosure of trade secrets could prove extremely costly in a commercial context. At the very least, redaction failures are embarrassing to those deemed responsible.

Information risks associated with digital redaction include:

1. Making bad decisions about the data to be redacted, the technical methods or process to be used and/or the suitability (primarily competency and diligence) of those tasked to do it;

ISO/IEC 27036
ISO/IEC 27037
ISO/IEC 27038
ISO/IEC 27039
ISO/IEC 27040
ISO/IEC 27041
ISO/IEC 27042
ISO/IEC 27043
ISO/IEC 27045
ISO/IEC 27046
ISO/IEC 27050
ISO/IEC 27070
ISO/IEC 27071
ISO/IEC 27090
ISO/IEC 27091
ISO/IEC 27099
ISO/IEC TS 27100
ISO/IEC 27102
ISO/IEC TR 27103
ISO/IEC TR 27109
ISO/IEC TS 27110
ISO/IEC TS 27115
ISO/IEC 27116
ISO/IEC 27400
ISO/IEC 27402
ISO/IEC 27403
ISO/IEC 27404
ISO/IEC TR 27550
ISO/IEC 27551
ISO/IEC 27553
ISO/IEC 27554
ISO/IEC 27555
ISO/IEC 27556
ISO/IEC 27557
ISO/IEC 27559
ISO/IEC TS 27560

2. Failing to identify correctly *all* the sensitive data that must be redacted (both the individual data items and the files);

3. Failing to render the redacted data totally unrecoverable, for example:

    - Using inappropriate or ineffective technical methods for redaction, such as crudely modifying rather than permanently deleting the sensitive data using methods that can be completely or partially reversed (for example simply reformatting or ~~overlaying~~ redacted text to *appear* invisible, or applying readily-reversed mechanistic transformations or tokenization of textual identifiers);

    - Accidentally leaving one or more copies of the sensitive data completely or partially unredacted (perhaps releasing multiple, independently and differently redacted versions of a sensitive document, enabling it to be reconstructed directly or by inference);

    - Partially deleting the sensitive data, leaving data remnants or sufficient information (such as the editing journal or cached copies) enabling the data to be restored from the redacted file;

    - Relying excessively on pixellation, blurring or similar methods of obfuscation to obscure parts of images (typically for personal privacy reasons), whereas deconvolution and other more or less advanced image manipulation/transformation techniques may restore enough of the original image to permit recognition;

    - Neglecting to redact sensitive metadata (*e.g*. in document properties or reviewer comments, GPS data on digital images, or alternate data streams);

4. Failing to distinguish all redacted from non-redacted data, consistently and accurately, such that recipients know unambiguously which parts are no longer original;

5. Excessive or inappropriate redaction, removing more than just the specific sensitive items that were supposed to have been redacted or doing so clumsily (which raises the prospect of having to justify redaction decisions and activities to a trustworthy intermediary or authority);

6. Inappropriately or inadvertently altering the meaning of the remaining data as a result of contextual issues (*e.g.* deleting selected data records may invalidate statistical analysis of the remainder), or by causing collateral damage to the file structure (such as file integrity issues and inappropriate formatting changes) during the redaction process;

7. Leaving sufficient data in the file to enable recipients to *infer* sensitive information, perhaps in conjunction with other available information sources (*e.g.* replacing people's names with anonymous labels in a redacted file but separately disclosing the relationship between labels and names; disclosing anonymous statistical data on known small populations; disclosing the number of characters redacted, and perhaps even giving clues to the most likely characters by dint of their printed size; applying data mining, correlation and inference techniques to glean sensitive data from redacted or anonymized content);

8. Placing excessive reliance on redaction, believing it sufficient to keep sensitive data totally confidential under all circumstances whereas technical and process failures are possible and incidents sometimes occur in practice; conversely, placing zero reliance on redaction, believing it to be totally incapable of protecting sensitive information (these are governance and assurance risks);

9. Information security issues that are incidental or peripheral to the redaction process itself such as:

    - Sending the original files, redaction instructions, redacted content or indeed the redacted files to the wrong recipients;

    - Failing to secure information relating to the redaction process, such as the original files or detailed redaction instructions, while in transit, during processing and in storage (*e.g*. interception of sensitive content in clear on the network);

    - Accidentally disclosing unredacted versions of the file, whether at the same time and through the same disclosure mechanism or separately;

    - Deliberate disclosure or 'leakage' of unredacted versions of the file without permission or inappropriately (*e.g.* to Wikileaks);

    - Accidentally or deliberately disclosing the redacted information by some means other than by releasing the digital data (*e.g.* by releasing the redaction instructions, or being overheard discussing sensitive matters);

    - Damaging the integrity and/or availability of the original unredacted files (*e.g*. overwriting them with the redacted versions);

10. Use of redaction to conceal illegal or inappropriate activities;

11. Use of AI/ML/NLP to surmise the redacted content based on linguistic principles and the surrounding context, plus broader analysis of related materials;

12. Various other risks (the risk analysis implied here is *generic* and *not comprehensive*: it does not necessarily reflect any specific situation).

ISO/IEC 27561
ISO/IEC 27562
ISO/IEC TR 27563
ISO/IEC 27564
ISO/IEC 27565
ISO/IEC 27566
ISO/IEC 27568
ISO/IEC 27569
ISO/IEC TS 27570
ISO/IEC 27573
ISO/IEC 27701
ISO/IEC 27706
ISO 27799

*[Thanks to colleagues on CISSPforum for contributing to this list.]*

## Scope and purpose

The standard formally defines *redaction* as “permanent removal of information within a document” where *document* is formally defined as “recorded information which can be treated as a unit”. The definitions are important because, in other contexts and general use, these terms often mean other things ... and indeed later in the standard, redaction is expanded to include not just the removal of confidential content but also, if appropriate, indicating where content has been removed.

The standard “specifies characteristics of techniques for performing digital redaction on digital documents [... and ...] requirements for software redaction tools and methods of testing that digital redaction has been securely completed [... but ...] does not include the redaction of information from databases.” Databases qualify as ‘units of recorded information’ but redaction of databases is specifically excluded from the scope of the standard.

Even though this standard has a restricted scope, the risks it covers are significant and many of the associated controls are technically and procedurally complex. Like other ISO27k standards, it does not attempt to cover all the vagaries of the redaction process in great detail but provides sound if rather generic and high-level guidance.

## Structure and content

After the usual preamble, scope and definitions, the bulk of the standard covers:

- An introduction to the general principles of digital redaction and anonymization of data;
- Redaction requirements - actually an overview of the redaction process;
- Redaction processes such as printing and physically redacting content, editing the original documents in various ways, dealing with metadata (such as document properties and change records) and, in the case of ‘enhanced’ redaction, considering the broader context as well as the specific content (*e.g.* the possibility of guessing, inferring or reconstructing redacted content from other content in redacted files, or by using other sources);
- Keeping records and notes in order to be able to explain or justify redaction decisions and actions;
- Software redaction tools - a core set of functional requirements;
- Redaction testing - five simple if basic ways to check whether the redaction has been successful; and
- An informative annex about redacting PDFs.

The title uses the keyword ‘specification’ which, in ISO-speak, implies a formal definition against which organisations may be independently audited and certified compliant.

## Status of the standard

The *first* edition was published in **2014** and confirmed unchanged in 2019.

## Personal comments

Whereas ISO specification standards normally use the key-word “shall” exclusively to indicate mandatory requirements, the DIS version also used “should” in places, providing guidance above and beyond the formal specifications. In practice, this makes the standard easier for users to understand and apply, but harder to audit and certify against, if indeed that was ever intended.

The standard doesn’t say much about the governance or overall management of the redaction process (*e.g.* identifying what has to be redacted, why, how and by whom, nor about analysing and treating the risks in a given redaction situation), nor on the security controls that ought to be applied to or associated with the process (*e.g.* to prevent the inappropriate release of unredacted content or explicit redaction instructions). There is room here for further implementation guidance.

< Previous standard ^ Up a level ^ Next standard >