EWG M2 Recommendation to the ICH Steering Committee
Electronic Standards for the Transfer of Regulatory Information (ESTRI)
File Integrity – MD5

Ver 1.0 10 June 2010

Title: File Integrity – MD5

Date: 10 June 2010

Background:
It is recognized that there is need for secure electronic regulatory information exchange among the three ICH regions. Critical to this secure information exchange is a method to ensure that the recipient has received exactly what the sender intended.

Recommendation:
It is recommended that a "checksum" be utilized to ensure this file integrity. A checksum or hash sum is a fixed-size datum computed from an arbitrary block of digital data for the purpose of detecting accidental errors that may have been introduced during its transmission or storage. The integrity of the data can be checked at any later time by recomputing the checksum and comparing it with the stored one. If the checksums do not match, the data was almost certainly altered (either intentionally or unintentionally).

Electronic submissions should contain checksums for each individual file transmitted. It is recommended that the MD5 Message-Digest Algorithm (MD5) should be used for this purpose. The use of a checksum for transmitting files provides a number of benefits including:

- The integrity of each file can be verified by comparing the checksum submitted with the file and the computed checksum.
- The checksum can be used to verify that the file has not been altered in the historical archive of the regulatory authority. This is particularly useful as files are migrated from one storage medium to another (e.g. when files are backed up to magnetic tape storage)

The exact implementation will be defined in ESTRI specifications for the exchange message (e.g. the ESTRI eCTD specification defines the precise way in which checksums are included).

Internal security and access control processes in the regulatory authority should maintain the integrity of the submitted files.

Conditions: None.

Remarks: MD5 is an open standard defined by the Internet Engineering Task Force (IETF) RFC 1321. ICH M2 recognizes there are documented flaws in MD5's design and that cryptographers have begun recommending use of other algorithms. However, MD5 is sufficient for the stated purpose of verifying file integrity.