
CESSDA's approach to bulk DDI study-level metadata validation and feedback

John Shepherdson*¹ and Matthew Morris*²

¹CESSDA – Norway

²CESSDA – Norway

Abstract

The CESSDA Data Catalogue (CDC) provides researchers with a single point of reference for the data holdings of CESSDA's Service Providers. It harvests study-level metadata in DDI XML format from numerous OAI-PMH endpoints. A high degree of metadata standardisation is required, in order to support sophisticated search and browsing techniques and provide researchers with relevant results.

The CESSDA Metadata Validator (CMV) has been developed to allow both data publishers and consumers to check metadata quality against published standards. DDI Profiles (formal, machine-actionable documents that specify additional constraints on the content of a DDI XML document, over and above those specified by the document's associated XSD schema) are used to define the standards that must be met by study-level metadata. The constraints are assigned to validation gates that build on one another and thus allow different levels of compliance to be specified and validated.

CMV has been incorporated into CDC and is used to check harvested records for compliance with specific DDI profile/quality gate combinations. Non-compliant records may be automatically excluded from the catalogue, or just flagged as such. A dashboard is available so publishers can view and locate any constraint violations flagged against their metadata records.

*Speaker