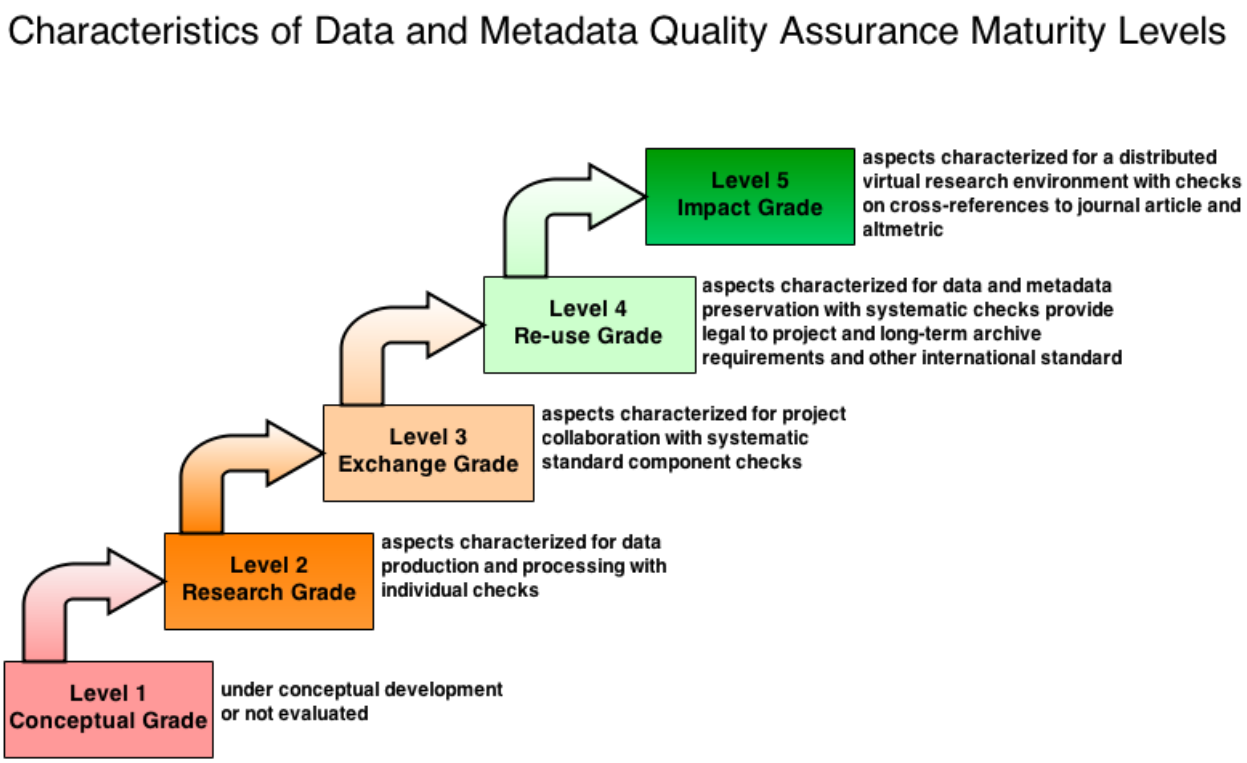
Quality Maturity Matrix (Version 03.04.2014)



Different criteria are defined, which are subdivided into aspects. For every aspect the 5 maturity quality levels are defined.

The Criteria are (Version 03.04.2014):

Consistency

- Data Organisation
- Versioning (inherent persistency of data)
- Controlled Vocabularies (CV)
- Data Object
- Data-Metadata Consistency
- Data internal Consistency
- Data temporal Consistency if applicable
- Data spatial Consistency if applicable
- Data thematic Consistency if applicable

Completeness

- Availability of Data
- Availability of Metadata

- Persistency of Data

Accessibility

- Technical Data Access (Identifier)
- Metadata Access
- Access Constraints
- Preservation (retain) - Persistency of Access

Provenance

- Provenance - Metadata
- Technical Provenance: PID (Lineage)

Accuracy

- Plausibility
- Statistical Anomalies

Usability

- Data Suitability
- Understandability/Use Constraints Data Object
- Operability/Use Constraints Discovery, Access
- Operability/Use Constraints Data Services
- Attractiveness
- Data Annotation (by other scientists)
- Citation Recommendation

State of the Quality Assessment System

The Quality Assessment System is currently developed. The information on this page will be updated as progress is made. Comments and suggestions are welcome.

Level 1 Conceptual Grade

Characteristics: under conceptual development or not evaluated

Consistency: conceptional development

- **Controlled Vocabularies (CV)**: simple or CV project requirements documented in data management plan

Usability: conceptual development

- **Attractiveness**: scientific questions

Level 2 Research Grade

Characteristics: aspects characterized for data production and processing with individual checks

Consistency:

- **Data Organisation:** informal data organisation
- **Versioning (inherent persistency of data)**: informal versioning, data might be overwritten
- **Controlled Vocabularies (CV):** informal CV
- **Data Object:** informal file names and extensions
- **Data-Metadata Consistency:** basic metadata are correct
- **Data internal Consistency:** missing values are indicated e.g. with fill values

Completeness:

- **Availability of Data:** data is in production, available for production group
- **Availability of Metadata:** basic documentation by researcher
- **Persistence of Data:** data may be deleted or overwritten

Accessibility:

- **Technical Data Access (Identifier):** accessible by name for production group
- **Metadata Access:**accessible for production group
- **Access Constraints:** production group

Provenance:

- **Provenance - Metadata:** who (creator), what (names) + identifier + data life-cycle unsystematically documented
- **Technical Provenance: PID (Lineage):** consequent usage of identifier e.g. file names or more sophisticated like PIDs

Accuracy:

- **Plausibility:** documented procedure about methodological and technical sources of errors and deviation/inaccuracy
- **Statistical Anomalies:** documented procedure about rough errors (outliers, missing data)

Usability:

- **Data Suitability:** feasibility of converting data metadata into project required versions
- **Understandability/Use Constraints Data Object:** usable for few scientists: format, variable names

- **Operability/Use Constraints Discovery, Access:** discovery and access needs additional knowledge
- **Operability/Use Constraints Data Services:** basic user documentation on data services
- **Attractiveness:** scientific questions
- **Data Annotation (by other scientists):** personal contact
- **Citation Recommendation:** on request

Level 3 Exchange Grade

Characteristics: aspects characterized for project collaboration with systematic standard component checks

Consistency:

- **Data Organisation**: feasibility of extraction and regridding
- **Versioning (inherent persistency of data):** systematic versioning, no data overwritten
- **Controlled Vocabularies (CV):** formal CV of standard components are correct
- **Data Object:** size and checksum of standard components are correct, file names, extensions and format are correct
- **Data-Metadata Consistency:** standard components to a documented procedure are correct
- **Data internal Consistency:** missing values are indicated e.g. with fill values

Completeness:

- **Availability of Data:** data entities are available, not complete, available for project members
- **Availability of Metadata:** extended metadata is available + checksum + data reference
- **Persistence of Data:** data may be deleted but not overwritten

Accessibility:

- **Technical Data Access (Identifier):** accessible by domain (data archive) specific identifier for production group and selected users
- **Metadata Access:** access constraints + extended metadata of basic components and checksums are accessible
- **Access Constraints:** access granted by production group
- **Preservation (retain) - Persistency of Access:** continuous access

Provenance:

- **Provenance - Metadata:** who (creator, contact), what (names) + identifier + datasets data life-cycle basically documented, e.g. in data headers
- **Technical Provenance: PID (Lineage):** identifier used and mapping (bijective) to objects documented e.g. PIDs in data header

Accuracy:

- **Plausibility:** documented procedure about methodological and technical sources of errors and deviation/inaccuracy
- **Statistical Anomalies:** documented procedure about rough errors (outliers, missing data) + documented procedure about systematic errors (changes in mean, variance and trends)

Usability:

- **Data Suitability:** suitable for project objective + almost all data metadata meet requirements of project
- **Understandability/Use Constraints Data Object:** usable for research community: format, standardized variable names
- **Operability/Use Constraints Discovery, Access:** discovery for research community: naming conventions, file size appropriate or size-reducing services in place
- **Operability/Use Constraints Data Services:** basic standardized documentation of data services available + user support
- **Attractiveness:** downloads>0 project use
- **Data Annotation (by other scientists):** point of contact available
- **Citation Recommendation:** documented

Level 4 Re-use Grade

Characteristics: aspects characterized for data and metadata preservation with systematic checks provide legal to project and long-term archive requirements and other international standard

Consistency: score 3 +

- **Data Organisation**: structured according to well-defined rules
- **Versioning (inherent persistency of data):** systematic versioning collection, no data overwritten, old versions stored
- **Controlled Vocabularies (CV):** formal CV of almost all data are correct
- **Data Object:** size and checksum of almost all data are correct + data format acceptable self-descriptive with format curation
- **Data-Metadata Consistency:** almost all components to a documented procedure are correct + data header and content are consistent
- **Data internal Consistency:** missing values are indicated e.g. with fill values + outliers concerning limits are documented
- **Data temporal Consistency if applicable:** temporal behaviour concerning limits is documented
- **Data spatial Consistency if applicable:** horizontal and vertical behaviour concerning limits is documented
- **Data thematic Consistency if applicable:** scientific consisteny among multiple data sets and their relationships is documented

Completeness:

- **Availability of Data:** data entities are available and complete (dynamic datasets - data stream not affected) number of data sets (aggregation) are correct
- **Availability of Metadata:** standard metadata is available + checksum + citation metadata with PID
- **Persistence of Data:** reused data are persistent, as long as long-term archive exists or registration of persistent identifier requires, minimum 10 years (see rules of good scientific practice)

Accessibility:

- **Technical Data Access (Identifier):** accessible by persistent identifier (PID) registrated with resolving to data access for reusers + full recovery (backup)
- **Metadata Access:** access constraints + standard metadata format and checksum accessible + full recovery (backup)
- **Access Constraints:** reusers - as open as possible within the framework of the legal possibilities and Privacy Policy
- **Preservation (retain) - Persistency of Access:** as long as long-term archive exists or registration of persistent identifier requires minimum 10 years see rules of good scientific practice + maintenance and updates of access services

Provenance:

- **Provenance - Metadata:** data type + Scientific Quality Assurance (approval + review) + who(creator, contact, publisher), what (title), how (method) + what for search and discovery + detailed description of data production steps available
- **Technical Provenance: PID (Lineage):** PID provenance access supported with persistent objects

Accuracy:

- **Plausibility:** documented procedure about methodological and technical sources of errors and deviation/inaccuracy + documented procedure with validation against independent data
- **Statistical Anomalies:** documented procedure about rough errors (outliers, missing data) + documented procedure about systematic (changes in mean, variance and trends) errors + documented procedure about random errors

Usability:

- **Data Suitability:** documentation of data analysis e.g. diagnostics on structure: phenomena + regional structure etc.
- **Understandability/Use Constraints Data Object:** self-describing data objects, fully machine-readable
- **Operability/Use Constraints Discovery, Access:** discovery for research community: naming conventions, file size appropriate or size-reducing services in place + user-friendliness of portals and download services and support team
- **Operability/Use Constraints Data Services:** full standardized documentation of data services available + user support

- **Attractiveness:** downlaods>0 scientific and commercial use
- **Data Annotation (by other scientists):** user annotations or forums supported
- **Citation Recommendation:** standardized and persistent (including data)

Level 5 Impact Grade

Characteristics: aspects characterized for a distributed virtual research environment with checks on cross-references to journal article and altmetric

Consistency: score 4 +

- **Data Organisation**: structured according to standardized rules
- **Versioning (inherent persistency of data):** documentation of not included newer versions
- **Controlled Vocabularies (CV):** CV standardized
- **Data Object:**continuous update/addition of external references, e.g. scientific publications + consistent to external scientific objects and up-to-date
- **Data-Metadata Consistency:** external metadata and data are correct with continuous update of external metadata
- **Data internal Consistency:** discussion in journal articles
- **Data temporal Consistency if applicable:** discussion in journal articles
- **Data spatial Consistency if applicable:** discussion in journal articles
- **Data thematic Consistency if applicable:** discussion in journal articles

Completeness: score 4 +

- **Availability of Data:** score 4
- **Availability of Metadata:** PID Data Description Document with cross references + annotations and other sources of feedback continuously updated
- **Persistence of Data:** score 4

Accessibility: score 4 +

- **Technical Data Access (Identifier):** accessibility within other data infrastructures including cross references
- **Metadata Access:** score 4
- **Access Constraints:** exchange of access constraints with external data infrastructures
- **Preservation (retain) - Persistency of Access:** modernisation of access services and accessibility within different data infrastructures

Provenance: score 4 +

- **Provenance - Metadata:** who (re-use with citation) + cross references + standard level system + data life-cycle available including internal and external objects e.g. software, articles
- **Technical Provenance: PID (Lineage):** external PID references supported + provenance chain

Accuracy: score 4 +

- **Plausibility:** references to evaluation results (data) and methods
- **Statistical Anomalies:** score 4

Usability: score 4 +

- **Data Suitability:** documented in journal article
- **Understandability/Use Constraints Data Object:** references to sources
- **Operability/Use Constraints Discovery, Access:** other scientific objects discoverable and accessible
- **Operability/Use Constraints Data Services:** information on external sources of information
- **Attractiveness:** reuse of data in scientific publications with citations
- **Data Annotation (by other scientists):** user annotation service in place
- **Citation Recommendation:** relations between data versions available