There are a number of dimensions to data quality – See extracts and links to two papers below.

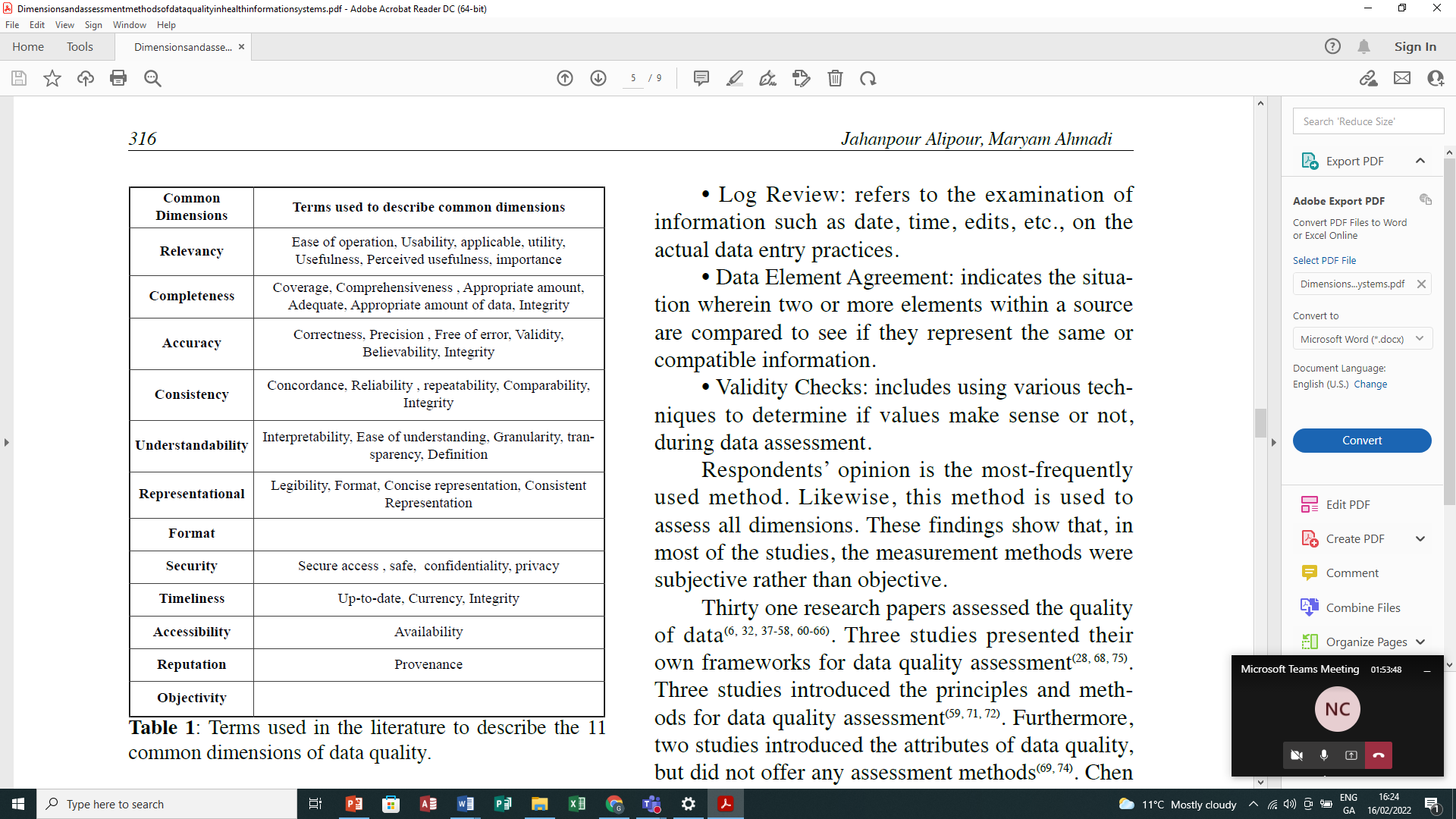
Paper 1

**Dimensions and assessment methods of data quality in health information**

**Systems**

Article *in* Acta Medica Mediterranea · March 2017

Authors - JAHANPOUR ALIPOUR1, 2, MARYAM AHMADI3, 4,\* 1 PhD Candidate of Health Information Management, Department of Health Information Management, School of Health Management and Medical Information Sciences, Iran University of Medical Sciences, Tehran, Iran - 2 Instructor, Department of Health Information Technology, Paramedical School, and member of Health Promotion Research Center, Zahedan University of Medical Sciences, Zahedan, IR Iran - 3 Professor of Health Information Management Department, School of Management and Medical Information Sciences, Iran University of Medical Sciences, Tehran, Iran - 4 Health management and economics research center, Iran University of medical sciences, Tehran, Iran



Paper 2

**Discovering Data Quality Problems**

Zhang, R., Indulska, M. & Sadiq, S. Discovering Data Quality Problems. *Bus Inf Syst Eng* **61,**575–593 (2019). https://doi.org/10.1007/s12599-019-00608-0

https://link-springer-com.elib.tcd.ie/article/10.1007/s12599-019-00608-0

|  |  |
| --- | --- |
| Dimension | Description |
| *Completeness* | |
| Completeness of mandatory attributes | The attributes which are mandatory for a complete representation of a real-world entity must contain values and cannot be null |
| Completeness of optional attributes | Optional attributes should not contain invalid null values |
| *Accuracy* | |
| Precision | Attribute values should be accurate as per linguistics and granularity |
| *Validity* | |
| Business rule compliance | Data should comply with business rules |
| Meta-data compliance | Data should comply with its meta-data |
| *Consistency* | |
| Uniqueness | The data is uniquely identifiable |
| Non-redundancy | The data is recorded in exactly one place |
| Semantic consistency | Data is semantically consistent |
| Value consistency | Data values are consistent and do not provide conflicting or heterogeneous instances |
| Format consistency | Data formats are consistently used |
| Referential integrity | Data relationships are represented through referential integrity rules |

**Notes**

OpenEHR is seen as a gold standard dictionary of definitions of electronic record fragments. These fragments can be communicated from system to system once the definitions are available.

There are a number of quality dimensions that require certain information to be captured in health information systems.

This project is going to analyse the archetype definitions to explore some of these quality dimensions.

Suggested items to extract from the OpenEHR Clinic Knowledge Manager (CKM)

|  |  |  |
| --- | --- | --- |
| Extract from OpenEHR | Rational | Related Quality Dimension |
| The data types of the contents of the archetype | Structured data is far easier to analyse for machines | Understandability |
| The number of time-related entries in the archetypes | Time and granularity of time are important in a number of use cases e.g.  To track disease progression  Medication management  To track impact of interventions  Medico legal reasons when trying to determine order of recording. | Timeliness |
| The number of languages used in archetypes |  | Format |
| Source of archetype | Provenance | Reputation |
| Coding system used | Coding systems are used to avoid confusion in interpretation of data | Representational |