Data quality in healthcare

23 janvier 2023

Module 3: Qualité des données - Majeure Santé EPITA

Dr Damien Leprovost – AP-HP / LIMICS

ASSISTANCE PUBLIQUE HÔPITAUX DE PARIS

Creative Commons BY-SA 4.0 License

ASSISTANCE PUBLIQUE HÔPITAUX DE PARIS

Course outline

Introduction to data quality challenges in healthcare

I. Fundamental notions
II. Why it's important

Data quality assessment framework

III. Quality dimensions
IV. Assessment orchestration

Data quality control procedures and tools

V. Framework content
VI. Data quality assessment tool

Interoperability and data quality

VII. Interoperability in healthcare
VIII. Existing solutions

Syllabus

- Define data quality
- Describe what drives data quality
- Explain importance of data quality for healthcare

- Highlight the notion of quality dimensions
- Present the quality framework

- Develop the strategy of quality assessment
- Express the assessment tool features

- Introduce the challenge of interoperability
- Explore existing solutions

I. Fundamental notions

What are we talking about?

Data In Healthcare

- Nowadays, data is central in healthcare
- The volume of clinical data is growing exponentially

Move from paper to electronic healthcare systems

Higher volume and variety of collected data

Data In Healthcare

Regarding healthcare, what data do you think of?

Where do these data come from?

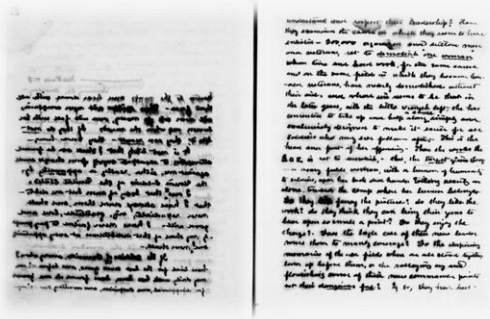
Various Healthcare Data Sources

Electronic Health Record

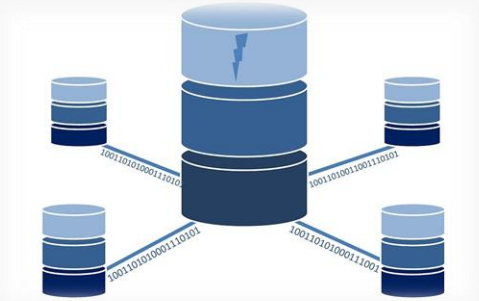
Providers' inputs

Device data

Test results



Document OCR



External databases

The purpose of healthcare data

- Document patient status
- Communicate care between providers
- Communicate care plans to patients

- Identify billable events
- Equipment and supply inventories
- Staffing decisions

- Support research projects
- Provide real life cohorts

The notion of Quality

- "Quality" is a wide-used term
- We are dealing every day with quality requirement
- But what defines:
  - The quality of a personal car?
  - The quality of a customer support?
  - The quality of a video?
  - The quality of this present course?
- What is the bottom line?

The notion of Quality

What means quality for you?

What does it imply?

The notion of Quality

- Quality can be boiled down to suitability, fit-to-purpose

- 2 basics aspects:
  - Validation
  - Verification

- 2 main outcomes:
  - Confidence
  - Repeatability

Data quality for healthcare

- Data quality **does not** aim to produce perfect data

- Data sufficiently accurate, timely and consistent
  - ➢ to make **appropriate and reliable decisions**

II. Why it's important

Challenges and risks

Quality drives quality

- The quality of decisions and plans will only be as good as the data one base them on

Treatment decisions require accurate and complete information to ensure quality care and the best possible outcomes

Billing, documentation of care and resources planification require a faithful representation of care activity

Scientific studies need robust data to defend reproducible findings to peers

Quality drives quality

- The quality of decisions and plans will only be as good as the data one base them on

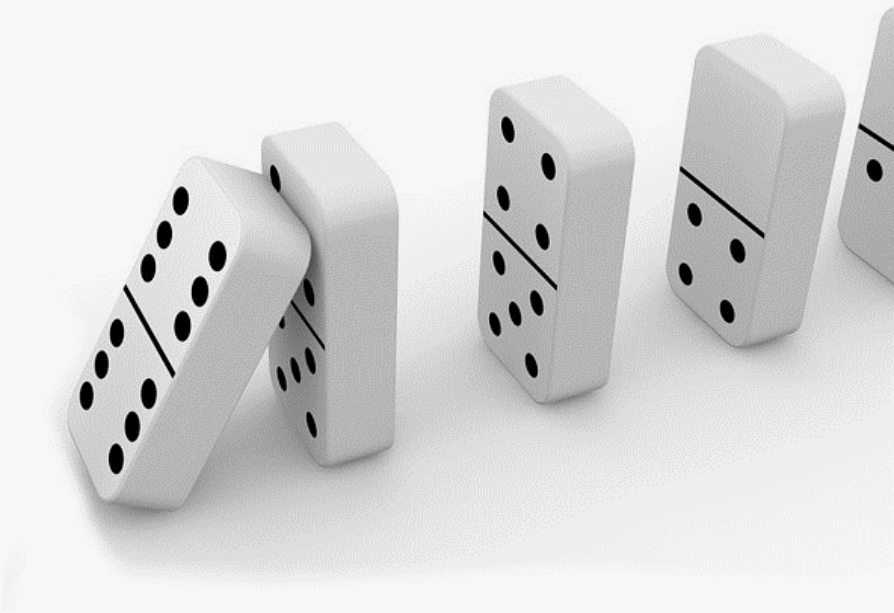
- Incorrect or incomplete data about patient
- Possible life-threatening decisions

- Billing incidents, waste of money
- Incorrect supply inventories, bad equipment decisions

- Flawed hypothesis and models
- Unusable outcomes

An implied quality

- Most users (and algorithms!) doesn't question the quality of the data they use
- Data unsuitable for use may lead to poor quality of service
- This must be guaranteed **upstream**



But a possible virtuous circle

- On the other hand, data users can have greater confidence in the data that they obtain when a systematic approach is taken to assessing the quality of that data.

Aims of data quality assessment

- Facilitation of better-informed strategic planning for healthcare
- Provision of safer, better-quality care for patients
- Better informed policies
- Greater contributions to research
- Improved population health

III. Quality Dimensions

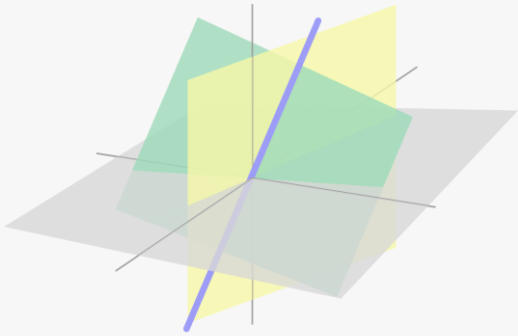
The main features of data quality

A multidimensional space

- Data quality is modelized as a **multidimensional** construct

- Each data quality dimension captures a particular **measurable** aspect of data quality

> The data quality dimensions represent the views, benchmarks or measures for data quality issues that can be understood, analyzed and resolved.



Dimensions aren't set in stone

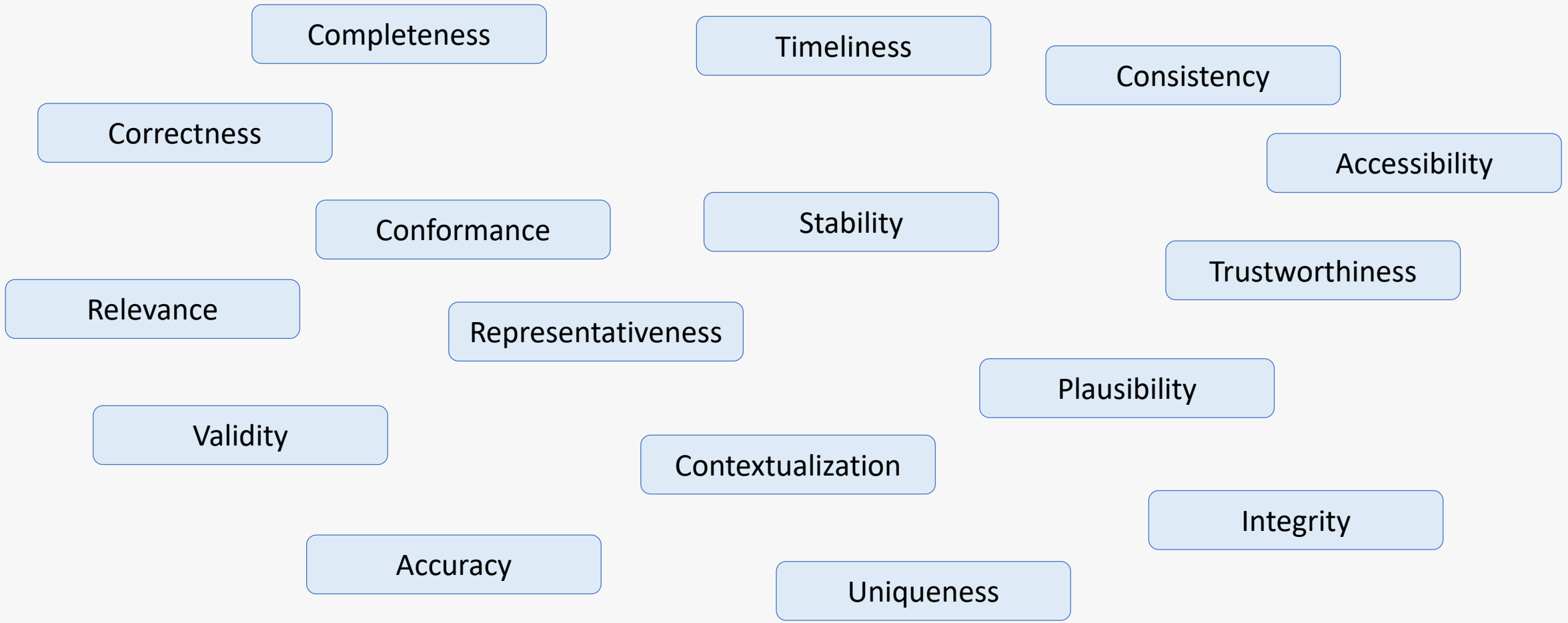
- They are many frequently used dimensions, but:
  - no consensus
  - no standardized set of dimensions
- Implementations commonly defines from 3 to 9 dimensions.
- Many of proposed dimensions aims to address similar data quality features.

Data Quality dimensions

What can be a monodimensional criterion?

What does it measure?

Some dimensions used in Data Quality



Common example: Completeness

- Definition:
  - The required data are stored in database
  - All related variables have values

- Examples:
  - The patient has a birthdate
  - The number of test results match the number of patient tested

- Purpose:
  - detect missing data and/or missing data values.

Common example: Consistency

- Definition:
  - The data values match their structural and syntactic constraints
  - All the values of a same variable match the same format and specifications

- Examples:
  - Patient admission time is recorded as a DATETIME with `YYYY-MM-DD hh:mm:ss` values
  - Insulin levels are stored as FLOAT expressed as µIU/L

- Purpose:
  - detect data which will not be usable

Common example: Timeliness

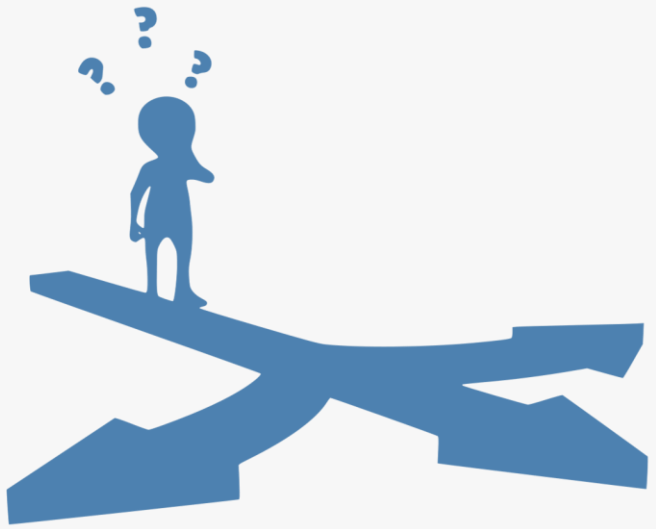
- Definition:
  - The data are up-to-date for their use
  - The data are available on time

- Examples:
  - Patient current care unit reflects the real situation of the patient
  - Yesterday's Covid-19 test result data are the aggregation of results from all test units in the hospital

- Purpose:
  - detect temporal inadequacy between alimentation flow and data usage

Dimensions of Data Quality

- Each dimension measure an **independent feature** of the data
- There's no fixed dimensions
- Dimensions must be chosen wisely, in accordance with:
  - The usages
  - The context
  - The resources



Example: The Dimensions of the Kahn Framework

- The Kahn Framework
  - M. Kahn et al., A Harmonized Data Quality Assessment Terminology and Framework for the Secondary Use of Electronic Health Record Data, EGEMS, 2016 https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5051581/pdf/egems1244.pdf
- A 2016-proposed data quality framework for the secondary use of electronic health record data
- Secondary use: the use of data for purposes other than those for which it originally was collected
  - operational review
  - quality improvement
  - research analytics
- Assesses quality on three dimensions
  - ► **Conformance**
  - ► **Completeness**
  - ► **Plausibility**

IV. Assessment orchestration

Why a framework is needed

Remember: Quality is relative

- Data quality is a **relative measure**
- It is dependent on financial and human resources
- Quality and quantity of data must be balanced against available resources

Why a framework is needed

- Data collection is a resource intensive process

- Organisations should recognize this from the outset to ensure that the amount and quality of data collected is aligned with the needs of data users.

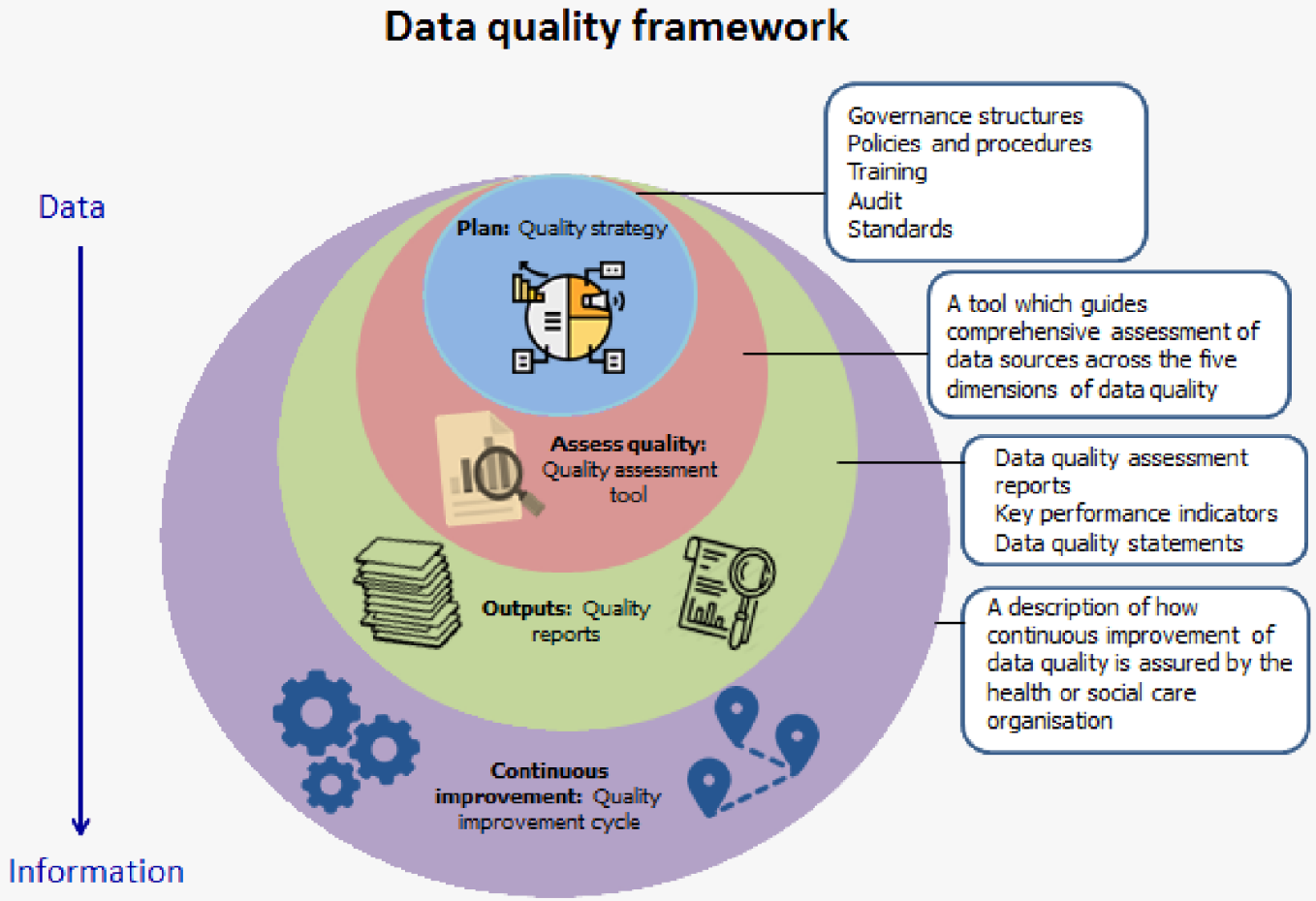
- It requires a thoughtful framework, to:
  - establish a baseline for data quality
  - identify any areas for improvement
  - assess the impact of any changes in practice, policies or procedures
  - report on improvements in data quality

Framework purposes

- A Data quality framework:
  - outlines the approach
  - provides the tools that can be used to:
    - assess data quality
    - document data quality
    - improve data quality

- The components of a data quality framework include:

Example: the HIQA Framework



source: Health Information and Quality Authority, Ireland – https://www.hiqa.ie

Data lifecycle

Capture
- When data enters either through data collection or acquisition from another organisation.

Submission
- The form and mechanisms by which data is submitted to an organisation, using standards, submission specifications and tools.

Processing
- The data is verified and cleaned prior to it being used. In some cases, data is transformed and aggregated for analyses.

Analysis
- Data are consumed and may undergo statistical analyses.

Dissemination
- Information produced is made available for data users.

source: Health Information and Quality Authority, Ireland – https://www.hiqa.ie

Data lifecycle and quality dimensions

- Applicability of a dimension of a particular data quality may vary:
  - at a particular stage
  - from one organization to another

- Carefully map out at what stage along the lifecycle a particular dimension of data quality should be applied

Stages \ Dimensions	Completeness	Consistency	Timeliness	Uniqueness
Capture			X	
Submission		X		
Processing	X			
Analysis				X
Dissemination	X	X	X	X

Example of lifecycle dimensions mapping

V. Framework content

Address data quality with guidelines

Keys components of the framework

A Data Quality Strategy

- A formalized approach to address data quality which sets out the activities needed to undertake.

A Data Quality Assessment Tool

- A set of criteria to comprehensively assess data sources across the defined dimensions of quality.

Reporting on data quality

- This may include internal or external data quality assessment reports, reporting on key performance indicators or metrics and producing "data quality statements".

A data quality improvement cycle

- A description of how continuous improvement of data quality is handle. It encompasses the processes and methodologies applied as part of their data quality improvement initiatives.

The Data Quality strategy

- A cluster of decisions to manage and continuously improve data quality
- Centred on organisational data quality goals
- Determine:
    - the data processes to improve
    - solutions to implement
    - people to engage
- Formalise the approaches for identifying, documenting and ensuring the implementation of data quality practices

Data Quality strategy implementation

- A data quality strategy can be a standalone document or can be incorporated within an overall information management strategy

> **Reminder:** Decisions within healthcare setting are only as good as the information on which they are based. Healthcare organisations have a responsibility to take the time and give the necessary commitment to ensure that high-quality data is established and maintained.

Data Quality strategy components

- Information on the following may be included as part of a data quality strategy:

Governance arrangements in relation to data quality	Data quality policies and procedures	Training, education and development programme in relation to data quality	Data quality audits	Standards for data quality
•The strategy should clearly define the governance arrangements, roles and responsibilities. •Data quality should be built in as an intrinsic part of the existing governance structures.	•Policies are written statements outlining an organisation's values in relation to the quality of their data. •Procedures describe the steps needed to achieve good quality data. •Policies and procedures should address each dimension of data quality.	• A skilled and competent workforce is essential in order to successfully implement a data quality strategy. • This programme should be informed by the findings of a training needs analysis. • Effective recruitment and workforce planning ensures that staff have the necessary competencies and skills.	•Data quality audits can be carried out internally or externally and should be undertaken on a regular basis. •The initial audit can be used as a benchmark for assessing improvements in data quality over time. •A quality improvement action plan should be produced, based on the findings of the audit.	•Adhering to standards facilitates better data quality and makes comparing information easier through supporting data sharing and interoperability. •The most commonly used standards include data dictionaries and terminology standards.

The Data assessment tool

- The data assessment tool provides a detailed set of criteria to assess its data sources across quality dimensions
- The dimensions are defined by the data quality strategy
- The dimensions and the assessment tool must be **stable over time**
  - Required condition for comparison and continuous improvement cycle

Reporting on data quality

- Report on the findings of data quality assessment
- Reporting on data quality can include:
  - internal or external data quality assessment reports
  - reporting on key performance indicators (KPIs)
  - the production of data quality statements

- Communicate to data users how ‘fit for purpose’ the data are

- Reporting on data quality can be undertaken internally and/or externally

Reporting: Data quality assessment reports

- Describe the quality of the data according to the dimensions defined by the data quality strategy
- Should include details on data quality assessments conducted to identify the strengths and weaknesses of the organisation's data
- Should be presented using a standardised reporting structure
  - facilitate readability and comparability

Reporting: Key performance indicators (KPIs)

- Specific and measurable elements
- Used to characterise the quality of data and report performance
- Designed to assess key aspects of structures, processes and outcomes of data quality

When reporting on data quality, it is important to establish and measure specific indicators in relation to each of the dimensions of data quality.

Reporting: Data quality statements

- A statement prepared to accompany significant published outputs
- Based on the findings of the data quality assessment undertaken
- Highlight the strengths and weaknesses of the data, assessed using the dimensions
- The content varies depending on the data being published, but is likely to include:
  - an overview of the data collection and its purpose
  - a description of the data source
  - a summary of the issues identified with the data
  - an overview of the quality of the data under each of the dimensions

Data quality improvement cycle

active **identification** of data issues

planning and execution of **assessment** processes

maintenance of a continuous data quality improvement cycle

implementation of quality **improvement** processes

evaluation of the quality improvement outcomes and processes

VI. Data quality assessment tool

Identify metrics to monitor

The data quality assessment tool

- The data assessment tool provides a detailed set of criteria to assess its data sources across quality dimensions
- Dimensions are not mutually exclusive, trade-offs between dimensions may be required at times throughout the data and information lifecycle

> The dimensions are not necessarily equally weighted across all organisations as the importance of each dimension may vary depending on the data source and context.

- The data quality assessment tool is organised in sections, each focusing on one key dimension of data quality.

Dimensions described here draws strongly from the CIHI's Data Quality Framework, CIHI's Information Quality Framework, CIHI's Data Source Assessment Tool, CIHI's Information Quality Plan, the European Statistical System Quality Assurance Framework and the United Nations Statistics Division National Quality Assurance Framework.

As depicted before, the characteristics and criteria included here are not an exhaustive list and may be amended by organisations as necessary to meet their needs and specifications.

Structure of the data quality assessment tool

Introduction to the dimension	An explanation of the dimension of data quality and the importance of assessing data quality under that dimension
The quality characteristics of the dimension	A description of the key quality characteristics, as defined by international best practice, that provide context for organisations on how data quality can be assessed to determine if it is 'fit for purpose'. Each characteristic is accompanied by a description which provides further explanation and clarity of understanding of the key characteristics Characteristic | Description
The assessment tool	Each characteristic is made up of a set of criteria, presented as key questions, which enable organisations to assess the degree of quality of their data. Each criteria is further explained by an accompanying details section, explaining in basic terms the practical measures that organisations must undertake in order to assess data quality. Characteristic | Criteria | Details | Assessment
Examples of assessment of that dimension	It is a good practice to provide examples of how national data collections currently assess the quality of their data under that dimension.

Example: Accuracy and reliability

Definition of the dimension
The accuracy of data refers to how closely the data correctly describes what it was designed to measure
Assessment Introduction
Healthcare organisations should have processes in place which allow them to produce and disseminate data which accurately portrays reality. For example, accurate data on the incidence of infectious diseases is vital when planning public health interventions such as vaccination programmes. Measuring accuracy enables to quantify the degree to which data correctly and consistently reflects the situation that it was designed to measure. The degree of accuracy of data has clear implications for how useful and meaningful the data will be for interpretation and further analysis. The accuracy of data is dependent upon the degree to which individual records received contain complete, valid information. Thus, the more complete the dataset is, the more likely the data is to be accurate. The accuracy of their data is quantified by the degree of difference between the estimate and the true value of the data quantify. Where inaccuracies are identified in the data, it should be acknowledged and addressed where possible.

Example: Accuracy and reliability

Characteristics of the dimension

Characteristic	Description
Coverage	The degree to which the data available to data users covers the population or event of interest. It is critical to understand the level of coverage of a population in order to make a statement about that population. Coverage can be assessed by defining the population in question and determining and monitoring coverage rates. Any limitations to the data as a result of coverage issues should be clearly explained.
Data capture and collection	The procedures that are in place to ensure that the data is captured in a usable format and can be prepared for submission. This is assessed by identifying what measures are in place to ensure that relationships with data providers are positive and that suitable processes are in place to facilitate efficient and effective data flows.

Example: Accuracy and reliability

Characteristics of the dimension

Characteristic	Description
Data processing	The transformation of data from the form in which it is received into another form that facilitates analysis. Processing can include validation and correction of the data. Assessment involves checking if processing has been documented in detail and if the processed data can be differentiated clearly from the original data received.
Completeness and validity	The degree to which individual variables are present within a data set. Incomplete and invalid data within a dataset impacts upon the fitness for use of that dataset. Assessment of completeness and validity is done by calculating and monitoring rates of valid, invalid, missing and outlier values.
Revisions to data	The extent to which the data is subject to revision or correction following new information or correction of errors in processing or estimation of data and the time frame in which revisions are produced.

Accuracy and reliability

Data quality assessment tool

Characteristic	Criteria	Details	Assessment
Coverage	Are details of the reference population explicitly stated in all information releases and is the coverage of the population quantified?	State if the reference population is documented and if coverage has been quantified. Where this has not occurred, provide details. The reference population is the population for which information should be available.	Yes / No / Partially
	Are significant coverage issues that may impact analysis and interpretation of data documented and made available to users?	Provide details of any significant issues in relation to coverage that have the potential to impact on data users' interpretation. Consider the impact of any identified coverage issues in terms of their impact at different levels of reporting. If coverage issues are not documented, detail why this is the case.	Yes / No / Not Applicable
	Are processes in place to identify and handle duplicate and potential duplicate records within the data?	Describe the processes that are in place to identify duplicates and calculate the duplication rate to give an indication of its occurrence. Provide details of how duplicates are removed once identified. If measures are not undertaken to identify and remove duplicates, explain why this is the case and provide an estimate of the number of duplicates that remain in the data.	Yes / No / Partially

Accuracy and reliability

Data quality assessment tool

Characteristic	Criteria	Details	Assessment
Data capture and collection	Are issues with the quality of data submitted, that have the potential to impact significantly on analysis and interpretation of that data, addressed and documented for users?	Provide details of identified data quality issues, including the action undertaken to resolve the issue, the time frame to resolve the issue and any potential impact that such issues have on interpretation and analysis of the data.	Yes / No / Not Applicable
Data processing	Are data validation processes applied consistently and are the processes documented for data users?	Provide details of the types of validation checks undertaken, for example, checks on the structure and integrity of the data, checks for missing data, checks that the data conforms to data source specifications and checks for any unusual data, that is, outliers.	Yes / No / Partially
Completeness and validity	Are rates of valid, invalid, missing and outlier values documented and updated routinely and reported with each data release?	Provide a data profile for data elements which include valid, invalid and missing value rates.	Yes / No / Partially
Revisions to data	Are revisions or corrections made to the data regularly analysed to ensure effective statistical use of same?	Provide details of the reasons for, timing of, and nature of revisions. Clearly identify data as preliminary or revised.	Yes / No

The Example of the Kahn Framework

- 2016-proposed data quality framework for the secondary use of electronic health record data
- Secondary use: the use of data for purposes other than those for which it originally was collected
  - operational review
  - quality improvement
  - research analytics

- Define three data quality dimensions, as categories
  - Conformance
  - Completeness
  - Plausibility
- Applied with two data quality assessment contexts
  - Verification
  - Validation

M. Kahn et al., A Harmonized Data Quality Assessment Terminology and Framework for the Secondary Use of Electronic Health Record Data, EGEMS, 2016
https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5051581/pdf/egems1244.pdf

The Example of the Kahn Framework

VERIFICATION		VALIDATION	
DEFINITION	**EXAMPLE**	**DEFINITION**	**EXAMPLE**
CONFORMANCE: DO DATA VALUES ADHERE TO SPECIFIED STANDARDS AND FORMATS?			
VALUE CONFORMANCE			
a. Data values conform to internal formatting constraints. b. Data values conform to allowable values or ranges.	a. Sex is only one ASCII character. b. Sex only has values "M," "F," or "U."	a. Data values conform to representational constraints based on external standards.	a. Values for primary language conform to ISO standards.
RELATIONAL CONFORMANCE			
a. Data values conform to relational constraints. b. Unique (key) data values are not duplicated. c. Changes to the data model or data model versioning.	a. Patient medical record number links to other tables as required. b. A medical record number is assigned to a single patient. c. Version 1 data does not include medical discharge hour.	a. Data values conform to relational constraints based on external standards.	a. Data values conform to all not-NULL requirements in a common multi-institutional data exchange format.
COMPUTATIONAL CONFORMANCE			
a. Computed values conform to computational or programming specifications.	a. Database- and hard-calculated Body Mass Index (BMI) values are identical.	a. Computed results based on published algorithms yield values that match validation values provided by external source.	a. Computed BMI percentiles yield identical values compared to test results and values provided by the CDC.

M. Kahn et al., A Harmonized Data Quality Assessment Terminology and Framework for the Secondary Use of Electronic Health Record Data, EGEMS, 2016

The Example of the Kahn Framework

VERIFICATION		VALIDATION	
DEFINITION	**EXAMPLE**	**DEFINITION**	**EXAMPLE**
COMPLETENESS: ARE DATA VALUES PRESENT?			
a. The absence of data values at a single moment in time agrees with local or common expectations. b. The absence of data values measured over time agrees with local or common expectations.	a. The encounter ID variable has missing values. b. Gender should not be null. c. Medical discharge time is missing for three consecutive days.	a. The absence of data values at a single moment in time agrees with trusted reference standards or external knowledge. b. The absence of data values measured over time agrees with trusted reference standards or external knowledge.	a. The current encounter ID variable is missing twice as many values as the institutionally validated database. b. A drop in ICD-9CM codes matches implementation of ICD-10CM
PLAUSIBILITY: ARE DATA VALUES BELIEVABLE?			
UNIQUENESS PLAUSIBILITY			
a. Data values that identify a single object are not duplicated.	a. Patients from a single institution do not have multiple medical record numbers.	a. Data values that identify a single object in an external source are not duplicated.	a. An institution's CMS facility identifier does not refer to a multiple institutions.

M. Kahn et al., A Harmonized Data Quality Assessment Terminology and Framework for the Secondary Use of Electronic Health Record Data, EGEMS, 2016

The Example of the Kahn Framework

VERIFICATION		VALIDATION	
DEFINITION	**EXAMPLE**	**DEFINITION**	**EXAMPLE**
ATEMPORAL PLAUSIBILITY			
a. Data values and distributions agree with an internal measurement or local knowledge. b. Data values and distributions for independent measurements of the same fact are in agreement. c. Logical constraints between values agree with local or common knowledge (includes "expected" missingness). d. Values of repeated measurement of the same fact show expected variability.	a. Height and weight values are positive. a. Counts of unique patients by diagnoses are as expected a. Distribution of encounters per patient or medications per encounter distributions are as expected b. Serum glucose measurement is similar to finger stick glucose measurement. b. Oral and axillary temperatures are similar. c. Sex values agree with sex-specific contexts (pregnancy, prostate cancer). d. Height values are similar when taken by two separate nurses within the same facility using the same equipment.	a. Data values and distributions (including subgroup distributions) agree with trusted reference standards or external knowledge. b. Similar values for identical measurements are obtained from two independent databases representing the same observations with equal credibility. c. Two dependent databases (e.g., database 1 abstracted from database 2) yield similar values for identical variables.	a. HbA1c values from hospital and national reference lab are statistically similar under the same conditions. a. Distribution of patients with cardiovascular disease diagnoses are similar to CDC rates for the same age and sex groups a. Readmission rates by age groups for Medicare patients agree with CMS values b. Diabetes ICD-9CM and CPT codes are similar between two independent claims databases serving similar populations. c. Recorded date of birth is consistent between EHR data and registry data for the same patient.
TEMPORAL PLAUSIBILITY			
a. Observed or derived values conform to expected temporal properties. b. Sequences of values that represent state transitions conform to expected properties. c. Measures of data value density against a time-oriented denominator are expected *based on internal knowledge*.	a. Admission date occurs before discharge date. b. Date of an initial immunization precedes date of a booster immunization. c. Similar counts of patient observations between extraction-transformation-load cycles. c. Counts of emergency room visits by month shows expected spike during flu season. c. Medications per patient-day are as expected	a. Observed or derived values have similar temporal properties across one or more external comparators or gold standards. b. Sequences of values that represent state transitions are similar to external comparators or gold standards. c. Measures of data value density against a time-oriented denominator are expected *based on external knowledge*.	a. Length of stay by outpatient procedure types conforms to Medicare data for similar populations. b. Immunization sequences match the CDC recommendations. c. Counts of emergency room visits by month shows spike during flu season that are similar to local health department reports. c. Medications per patient-day matches claims data.

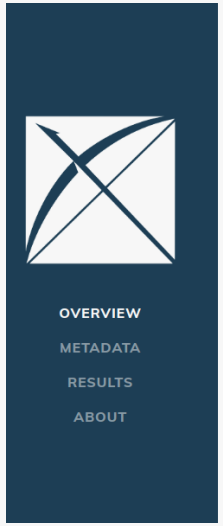
M. Kahn et al., A Harmonized Data Quality Assessment Terminology and Framework for the Secondary Use of Electronic Health Record Data, EGEMS, 2016

The Example of the OHDSI DQD

- OHDSI defines a Common Data Model (CDM) for the systematic analysis of disparate observational databases
  - The OHDSI OMOP CDM: https://www.ohdsi.org/data-standardization/the-common-data-model/
- Use of a CDM allows **sharable implementations** of assessment rules and tools
- Two notable contributions:
  - The ACHILLES Rules (Automated Characterization of Health Information at Large-scale Longitudinal Evidence Systems)
  - The Data Quality Dashboard

The Example of the OHDSI DQD

- Based on the Kahn Framework
- Runs data quality checks
- Evaluates the checks against some pre-specified threshold
- Communicates what was done in a visual interface



DATA QUALITY ASSESSMENT

SYNTHEA SYNTHETIC HEALTH DATABASE

Results generated at 2019-08-22 14:15:06 in 29 mins

	Verification				Validation				Total			
	Pass	Fail	Total	% Pass	Pass	Fail	Total	% Pass	Pass	Fail	Total	% Pass
Plausibility	159	21	180	88%	283	0	283	100%	442	21	463	95%
Conformance	637	34	671	95%	104	0	104	100%	741	34	775	96%
Completeness	369	17	386	96%	5	10	15	33%	374	27	401	93%
Total	1165	72	1237	94%	392	10	402	98%	1557	82	1639	**95%**

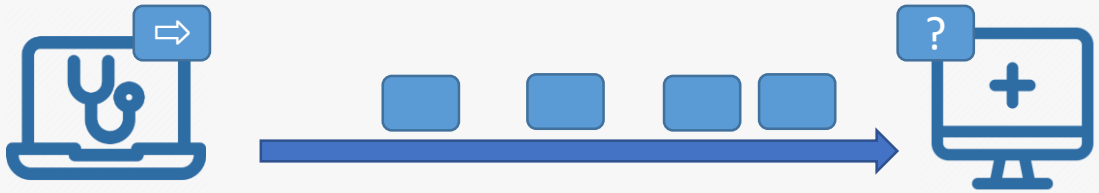
Live example: https://data.ohdsi.org/DataQualityDashboard/

VII. Interoperability in healthcare

Challenges and risks

Interoperability in healthcare

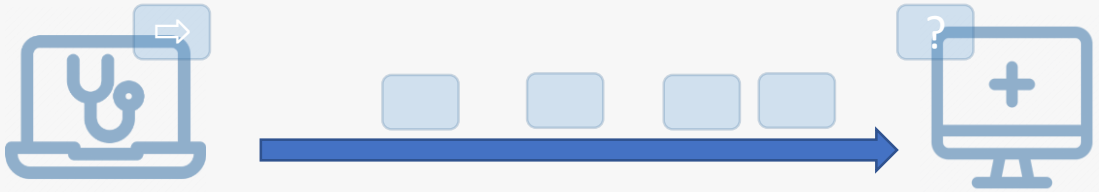
- The ability of different information technology systems and software applications to:
  - communicate
  - exchange data
  - use the information that has been exchanged



- Data exchange schema and standards should permit data to be shared regardless of the application or its vendor
- Interoperability means the ability of health information systems to work together within and across organizational boundaries

Strcutural Interoperability

- The ability of two (or more) systems to emit and receive data
- To perform a given task without the need for extra operator intervention



- Require agreement on protocols, APIs, security
- Domain independent

Semantic interoperability

- The ability of health IT systems to exchange and interpret information
- Both sender and recipient have data that means exactly the same thing

- Require agreement on codes and usages
- Specific to domain and concepts

Organisational interoperability

- The ability to transmit and use relevant data as part of the patient care
- Use new information in some way that differs from what they would have done without it, enabling to obtain different results and outcomes

- The human level on the interoperability
- Requires changes in workflow and in the way clinicians and clinical systems function at a fine level of detail

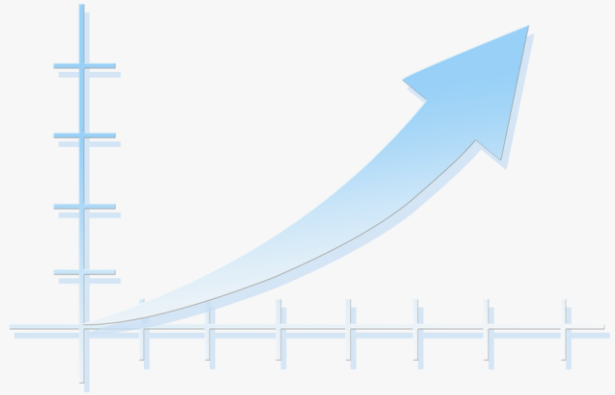
Interoperability in healthcare

- Clinical interoperability may require a major effort to re-engineer clinical workflows
- The more we understand all of the different aspects of interoperability, the less likely we are to underestimate the work required to make health systems interoperable.
- Technical, semantic, process and clinical interoperability are interdependent
  - all are needed to deliver significant business benefits

In a nutshell, interoperability aims to save an enormous amount of duplication, waste and errors.

An growing need for interoperability

- Easy access to patient record in the context of multi-care providers
- Explosion of the volumetry and diversity of health-related records
- Enhanced support for the management of chronic diseases
- Reduction of medical errors
- Limitation of the healthcare increasing costs



Barriers to interoperability in healthcare

What restrains healthcare interoperability?

Why isn’t it solved already?

Barriers to interoperability in healthcare

Domain complexity

Standardization problems

Overwhelming data

Legacy systems

Lack of skills and resources

Clinical ecosystem and market

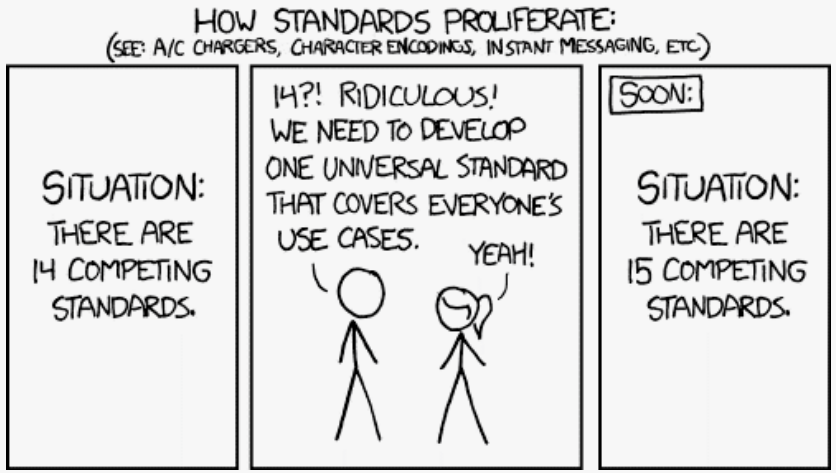
VIII. Existing solutions

Existing standards and terminologies

Interoperability initiatives in healthcare

- Over 40 different standard development organisations (SDO)
  - Define standards
  - Publish implementation guides and profiles

- Wide variety of data types
  - Vocabulary / Terminology standards
    - SNOMED-CT, LOINC, ICD
  - Content standards
    - HL7 CDA, OHDSI OMOP CDM
  - Transport standards
    - FHIR, DICOM
  - Privacy and Security standards
    - HIPAA rules, "GDPR"
  - Identifier standards
    - EMPI, MRN, ObjectID



xkcd – https://xkcd.com/927/ – CC NY-CA 2.5

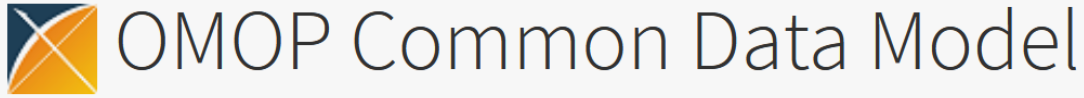
Focus on: HL7 FHIR

- Standard for exchanging healthcare information electronically
- FHIR resources describe exchangeable health data formats and elements
- API standardization
- Facilitating interoperable exchange with legacy standards

- First normative version is **FHIR Release 4**, published on Oct, 2019
- Connectathons events host development and testing workshops, as well as working group meetings.

Focus on: OHDSI OMOP CDM

- Standard for electronic health record observational database storage
- Describe data in a common format (model) and a common representation (terminologies, vocabularies, coding schemes)
- Designed for systematic analyses based on the CDM
- Facilitating interoperable exchange with legacy standards



- Current CDM version is **OMOP CDM v5.4**, published on Sept, 2021
- Yearly OHDSI Symposium hosts panels, workshops and tutorials.

Vocabularies & Terminologies

- Some useful vocabularies to know:

Name	Distribution	Domain	Type	Codes (approx)
ICD-10	Open (WHO)	Conditions, observations	Classification	12 000
SNOMED-CT	Licence required	Conditions, procedures, indications, drugs, devices, …	Terminology	311 000
LOINC	Free	Observations, measurements, documents, …	Terminology	95 000
MeDRA	Licence required	Conditions, procedure, indications, …	Classification	107 000
RxNorm	Open (US NLM)	Drugs (US)	Terminology	250 000
MeSH	Open (US NLM)	Indexing of journal articles	Vocabulary	41 000
CCAM	Open (FR HAS)	Procedures	Classification	7 600
ATC	Open (OMS)	Drugs	Classification	6 300

FAIR Data

- Humans increasingly rely on computational support to deal with data
  - Due to because the increase in volume, complexity, and creation speed of data
- The FAIR principles emphasize machine-actionability
  - the capacity of computational systems to find, access, interoperate, and reuse data with none or minimal human intervention

Finable

•Data and metadata should be easily findable by both humans and machines. This includes being assigned a unique identifier, described with rich metadata and indexed correctly.

Accessible

•The data not only needs to be findable, it also needs to be accessible; this could include everything from accessing immediately through to details of how to gain authentication Or authorization where required.

Interoperable

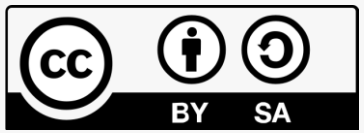
•Data should be interoperable, using a formal, accessible, shared, and broadly applicable language for knowledge representation. If data is siloed, then elements of it can be inaccessible or take a long time to access.

Reusable

•Ensuring data is reusable means that those who need to access it later on can do so easily and efficiently and with confidence that it is accurate) regardless of how long, or in what format it has been stored.

Course URL and license

Permalink : https://www.damien-leprovost.fr/enseignements/DataQualityHealthcare.2023.pdf



ASSISTANCE PUBLIQUE HÔPITAUX DE PARIS

References

- F. Dravis, Data Quality Strategy: A Step-by-Step Approach, ICIQ, 2004
  - http://mitiq.mit.edu/ICIQ/Documents/IQ%20Conference%202004/Papers/DQStrategy.pdf
- M. Kahn et al., A Harmonized Data Quality Assessment Terminology and Framework for the Secondary Use of Electronic Health Record Data, EGEMS, 2016
  - https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5051581/pdf/egems1244.pdf
- European Statistical System, Quality Assurance Framework of the European Statistical System, 2015
  - http://ec.europa.eu/eurostat/documents/64157/4392716/qaf_2012-en.pdf
- UN Statistics Division. Guidelines for the template for a generic national quality assurance framework, 2012
  - http://unstats.un.org/unsd/dnss/docs-nqaf/GUIDELINES%208%20Feb%202012.pdf
- Canadian Institute for Health Information, CIHI's Information Quality Plan, 2017
  - https://www.cihi.ca/sites/default/files/document/corporate-information-quality-plan-2016-en.pdf
- Health Information and Quality Authority (Ireland), Guidance on a data quality framework, Oct. 2018
  - https://www.hiqa.ie/reports-and-publications/health-information/guidance-data-quality-framework-health-and-social-care
- Healthcare Information and Management Systems Society, Definition of Interoperability, 2013
  - https://www.himss.org/sites/hde/files/d7/FileDownloads/HIMSS%20Interoperability%20Definition%20FINAL.pdf