



DATA VALIDATION

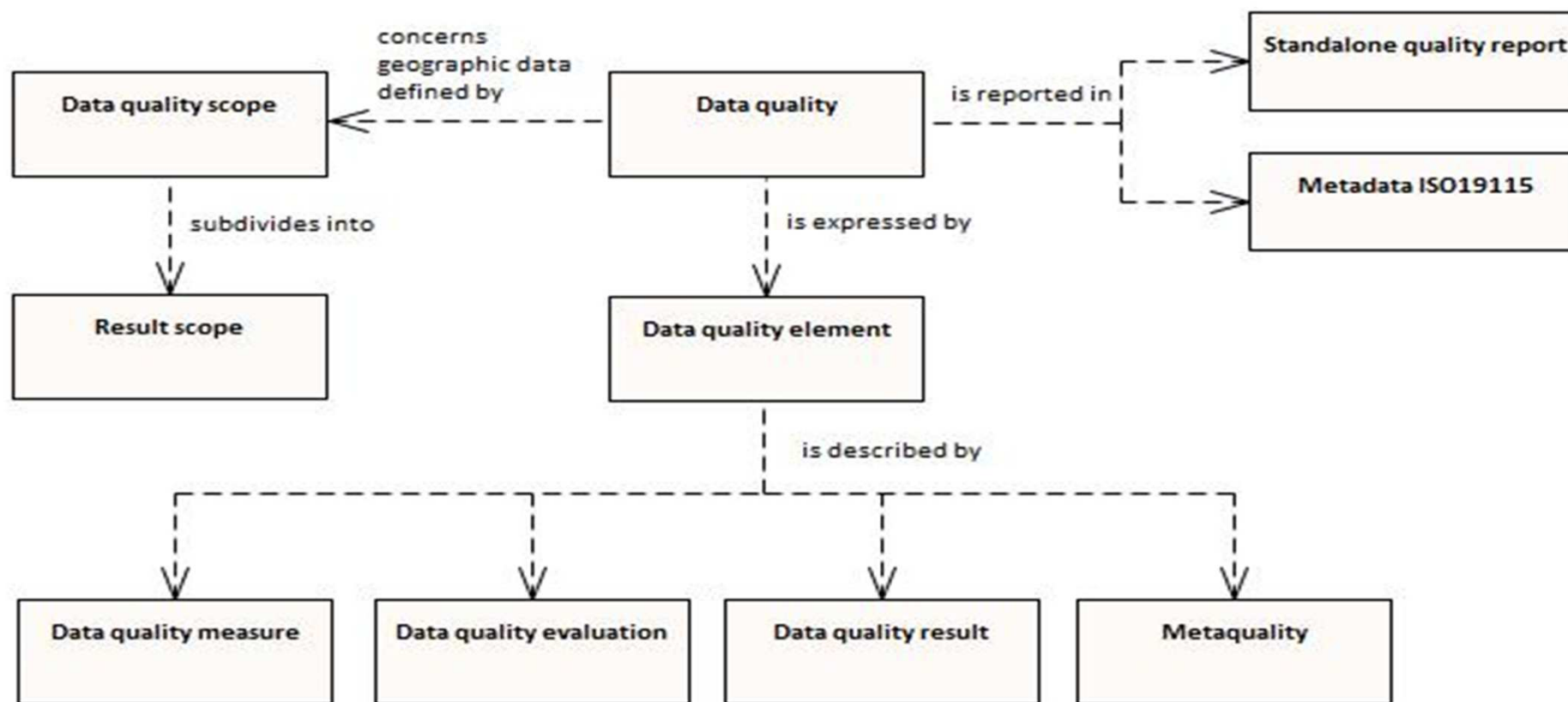
ISO principles



IHO

CONCEPT OF DATA QUALITY

International
Hydrographic
Organization



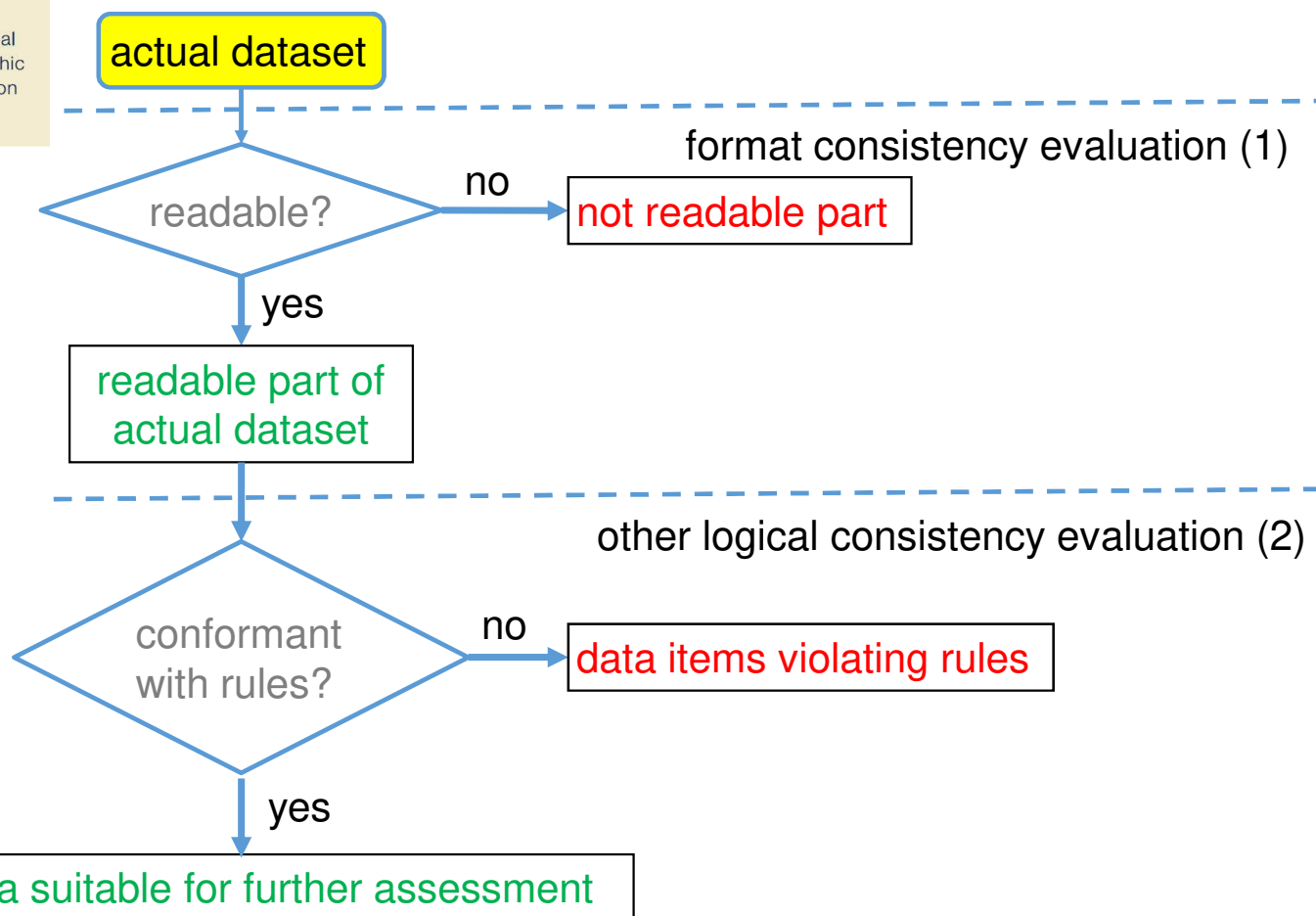
Data Validation ISO principles



IHO

ISO 19157 ORDERING IN DATA QUALITY EVALUATION

International
Hydrographic
Organization

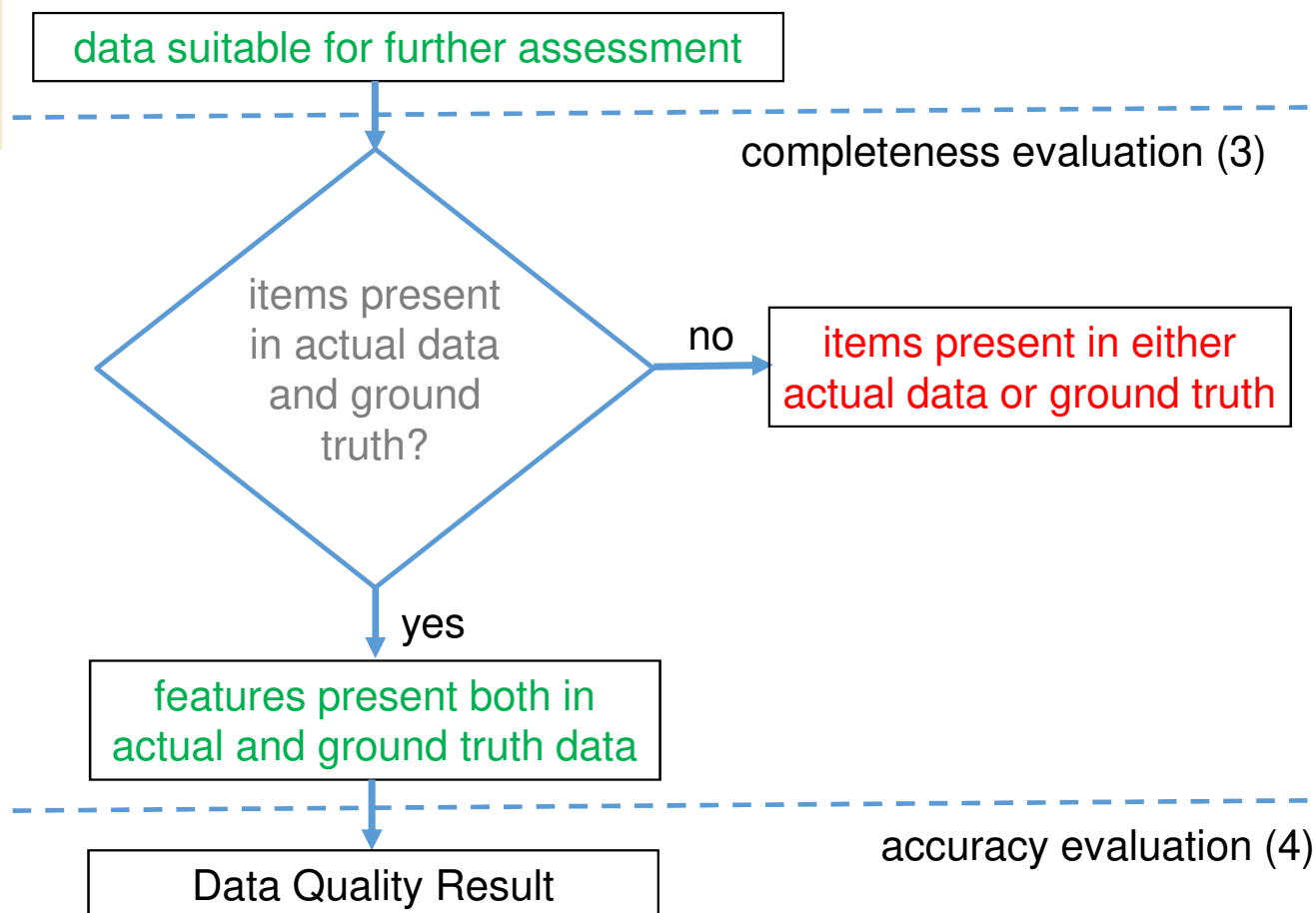




IHO

ISO 19157 ORDERING IN DATA QUALITY EVALUATION

International
Hydrographic
Organization





IHO

FORMAT CONSISTENCY

International
Hydrographic
Organization

- Format consistency – degree to which data is stored in accordance with the physical structure of the dataset
- Format consistency is described in **S-100 part 10** – Encoding formats
- S-100 does not mandate particular encoding formats so it is left to developers of product specifications to decide on **suitable encoding standards** and to document their chosen format. The issue of encoding information is complicated by the range of encoding standards that are available, which include but are not limited to: ISO/IEC8211, GML, XML, GeoTiff, HDF-5, JPEG2000.



IHO

International
Hydrographic
Organization

LOGICAL CONSISTENCY - DEFINITION

- Logical Consistency is defined as the degree of adherence to logical rules of data structure, attribution, and relationships (data structure can be conceptual, logical or physical). If these logical rules are documented elsewhere (for example in a data product specification) then the source should be referenced (for example in the data quality evaluation).



IHO

LOGICAL CONSISTENCY ITEMS

International
Hydrographic
Organization

- **conceptual** consistency – adherence to rules of the conceptual schema
- **domain** consistency – adherence of values to the value domains
- **topological** consistency – correctness of the explicitly encoded topological characteristics of a dataset



IHO

CONCEPTUAL CONSISTENCY

International
Hydrographic
Organization

- **S-100 part 1**, conceptual schema language. It provides description of:
 - classes
 - attributes
 - basic data types
 - primitive types
 - complex types
 - predefined derived types
 - enumerated types
 - codelist types
 - relationships and associations
 - composition and aggregation
 - stereo types
 - optional, conditional and mandatory attributes and associations
 - naming and name spaces
 - notes
 - packages



IHO

DOMAIN CONSISTENCY

International
Hydrographic
Organization

This is described in **S-100 Part 5** – Feature Catalogue.

This Part provides a standard framework for organizing and reporting the classification of real world phenomena in a set of geographic data. It defines the methodology for classification of the feature types and specifies how they are organized in a feature catalogue and presented to the users of a set of geographic data. This methodology is applicable to creating catalogues of feature types in previously uncatalogued domains and to revising existing feature catalogues to comply with standard practice. It applies to the cataloguing of feature types that are represented in digital form. Its principles can be extended to the cataloguing of other forms of geographic data.



IHO

TOPOLOGICAL CONSISTENCY

International
Hydrographic
Organization

- This is described in **S-100 Part 7** – Spatial Schema. It supports 0, 1, 2, and 2.5 dimensional spatial schemas and two levels of complexity – geometric primitives and geometric complexes.
- S-101 Validation Checks.xlsx lists a number of Topological checks.
- Inherited from S-58 Validation checks that apply to S-57 Topological Validation.
- Based on ISO 19125-1:2004 Geometry



IHO

International
Hydrographic
Organization

DEFINITIONS FOR ISO 19125-1: 2004 GEOMETRY

- *Polygon* - A Polygon has a geometric dimension of 2. It consists of a boundary and its interior, not just a boundary on its own. It is a simple planar surface defined by 1 exterior boundary and 0 or more interior boundaries. The geometry used by an S-57 Area feature is equivalent to a Polygon.
- *Polygon boundary* - A Polygon boundary has a geometric dimension of 1 and is equivalent to the outer and inner rings used by an S-57 Area feature
- *Line String* - A LineString is a Curve with linear interpolation between Points. A LineString has a geometric dimension of 1. It is composed of one or more segments – each segment is defined by a pair of points. The geometry used by an S-57 Line feature is equivalent to a LineString



IHO

DEFINITIONS FOR ISO 19125-1: 2004 GEOMETRY

International
Hydrographic
Organization

- *Line* - An ISO 19125-1:2004 line is a LineString with exactly 2 points. Note that the geometry used by an S-57 Line feature is equivalent to a LineString, not a line in ISO 19125-1:2004 terms. In this document the term Line refers to an S-57 Line feature or a LineString which can have more than two points.
- *Point* - Points have a geometric dimension of 0. The geometry used by an S-57 Point feature is equivalent to an ISO 19125-1:2004 point.
- *Reciprocal* – inversely related or opposite.

ISO 19125-1:2004	S-101
Polygon	Area feature geometry OR Area
Polygon boundary	Outer and inner rings
LineString	Line feature geometry OR Line
Point	Point feature geometry OR Point

Data Validation ISO principles



IHO

International
Hydrographic
Organization

GEOMETRIC OPERATOR RELATIONSHIPS

- In ISO 19125-1:2004 the dimensionally extended nine-intersection model (DE-9IM) defines 5 mutually exclusive geometric relationships between two objects (Polygons, LineStrings, and/or Points). One and only one relationship will be true for any two given objects:
 1. WITHIN
 2. CROSSES
 3. TOUCHES
 4. DISJOINT
 5. OVERLAPS



IHO

OTHER OPERATORS TO HELP DEFINE THE RELATIONSHIP

International
Hydrographic
Organization

1. CONTAINS

- the reciprocal of WITHIN
- within is the primary operator; however, if a is not within b then a may contain b so CONTAINS may be the unique relationship between the objects.

2. EQUAL

- a special case of WITHIN / CONTAINS.

3. INTERSECTS

- reciprocal of DISJOINT
- have at least one point in common

4. COVERS and is COVERED_BY

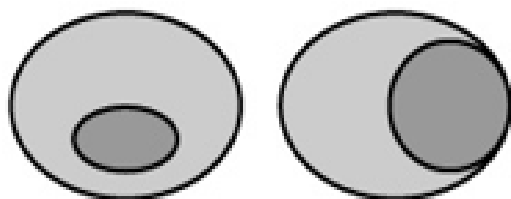
- reciprocal operators
- extends CONTAINS and WITHIN respectively

5. COINCIDENT



IHO

EXAMPLE WITHIN



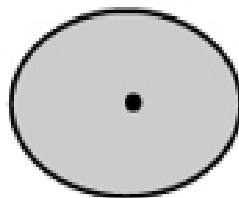
a)



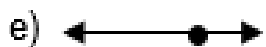
b)



c)



d)



e)

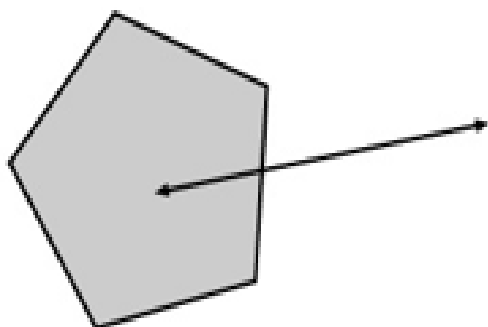
- a) Polygon / Polygon
- b) Polygon / LineString
- c) LineString / LineString
- d) Polygon / Point
- e) LineString / Point



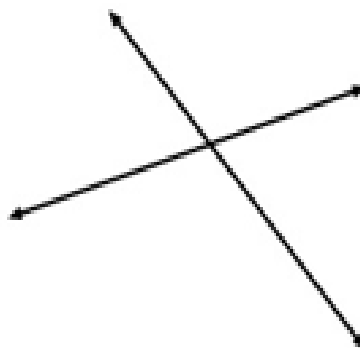
IHO

EXAMPLE CROSSES

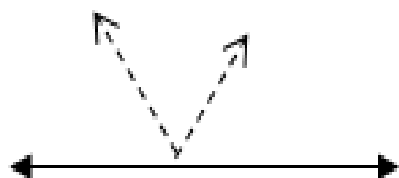
International
Hydrographic
Organization



a)



b)



c)

Note that example c) shows one solid line and one dashed line – their interiors intersect.

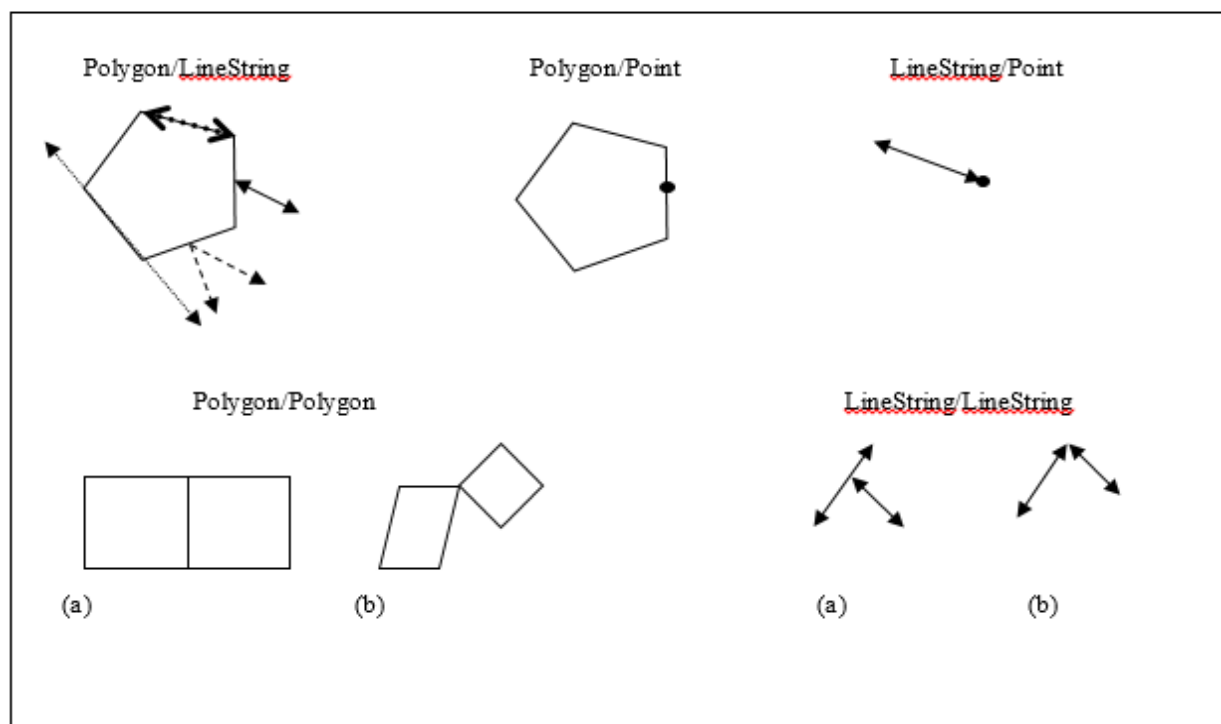
If any Line were split into two separate Line features at the intersection point then the relationship would be TOUCHES because a boundary would be involved.



IHO

EXAMPLE TOUCHES

International
Hydrographic
Organization



Note the Polygon touches Polygon example (a) is also a case where the Polygon boundaries are COINCIDENT.

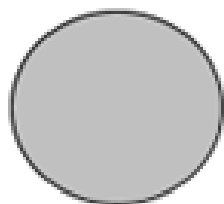
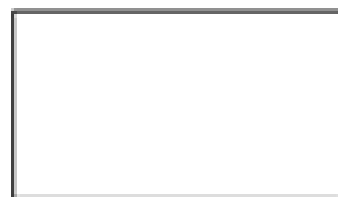
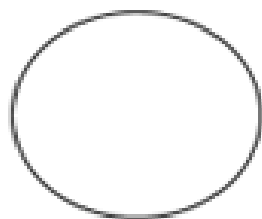
In the Polygon/LineString example two of the LineStrings that share a linear portion of the Polygon boundary are also COINCIDENT with the Polygon boundary



IHO

EXAMPLE DISJOINT

International
Hydrographic
Organization



(a)



(b)



(c)

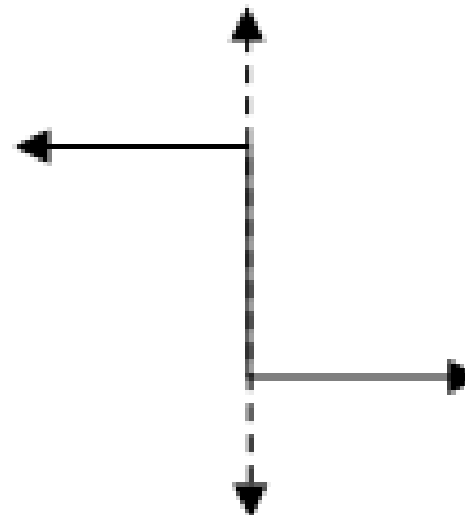
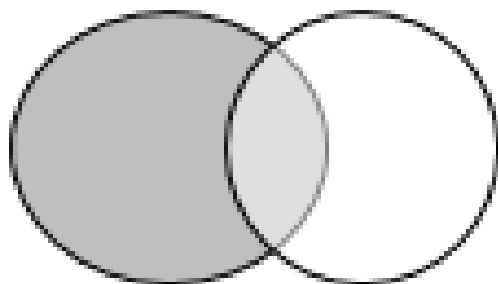
This translates to: Geometric object **a** is disjoint from Geometric Object **b** if the intersection of **a** and **b** is the empty set.



IHO

International
Hydrographic
Organization

EXAMPLE OVERLAPS



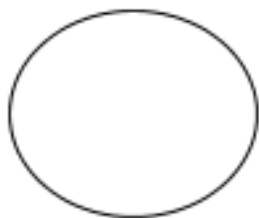
Note: Lines that OVERLAP are also COINCIDENT



IHO

International
Hydrographic
Organization

EXAMPLE EQUALS

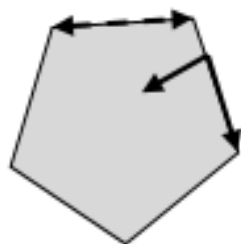


Geometric object **a** is spatially equal to geometric object **b**.

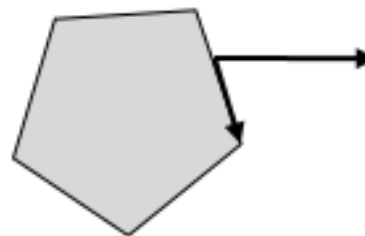


IHO

EXAMPLE COVERS AND IS COVERED BY



LineStrings
COVERED_BY
Polygon



LineString NOT
COVERED_BY
Polygon but
TOUCHES

Given two geometric objects,
a and **b**,
if **a** is COVERED_BY **b**
then **b** must cover **a**

No point of geometry **a** is
outside geometry **b**.

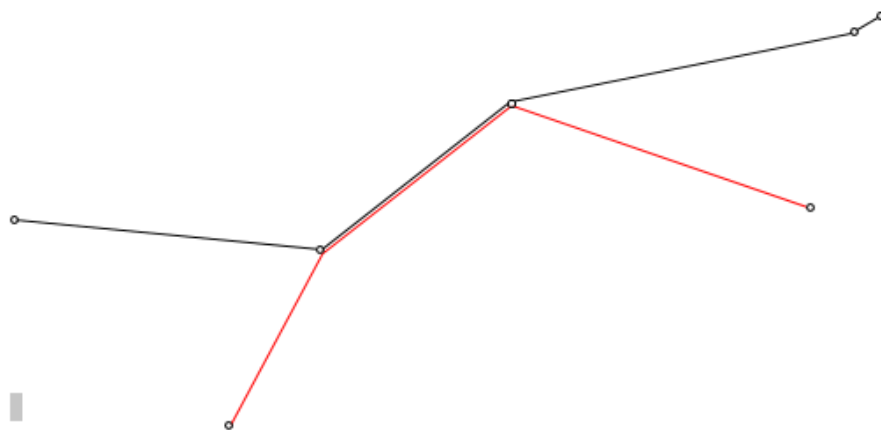
Note that the figure above on the left is an example of Lines that are COVERED_BY a polygon. The figure on the right is NOT an example of a Line that is covered by a Polygon – it is an example of a Line that TOUCHES a Polygon. In both cases the Lines are COINCIDENT with the Polygon boundary.



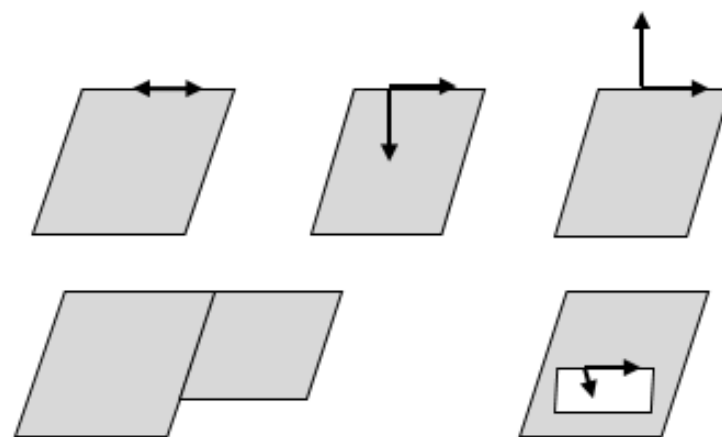
IHO

EXAMPLE COINCIDENT

International
Hydrographic
Organization



Example of two coincident lines.



Above are examples of objects
COINCIDENT with the boundary of a
Polygon. LineStrings following a portion
of a Polygon boundary or Polygons
sharing a boundaryportion.
*Note that by definition a Line can be
COINCIDENT with an interior boundary of
a Polygon.*



IHO

COMPLETENESS

International
Hydrographic
Organization

- **Completeness** is defined as the presence and absence of features, their attributes, and relationships. It consists of two data quality elements:
- **commission**, excess data present in a dataset;
- **omission**, data absent from a dataset.



IHO

ACCURACY

International
Hydrographic
Organization

- **Positional** accuracy is defined as the accuracy of the position of features within a spatial reference system. It consists of three data quality elements:
- **absolute** or **external** accuracy: closeness of reported coordinate values to values accepted or as being true;
- **relative** or **internal** accuracy: closeness of the relative positions of features in a dataset to their respective relative positions accepted as or being true;
- **gridded data** positional accuracy: closeness of gridded data spatial position values to values accepted as or being true.



IHO

METAQUALITY

International
Hydrographic
Organization

- Metaquality = information describing the *quality* of data quality
- Metaquality describes the quality of the data quality results in terms of defined characteristics
- Metaquality elements are a set of quantitative and qualitative statements about a quality evaluation and its result. The knowledge about the quality and the suitability of the evaluation method, the measure applied and the given result may be of the same importance as the result itself



IHO

DESCRIBING METAQUALITY

International
Hydrographic
Organization

- **Confidence** – trustworthiness of a data quality result
- **Representativity** – degree to which the sample used has produced a result which is representative of the data within the data quality scope
- **Homogeneity** – expected or tested uniformity of the results obtained for a data quality evaluation