

# Landupplýsingar og gæðastaðlar

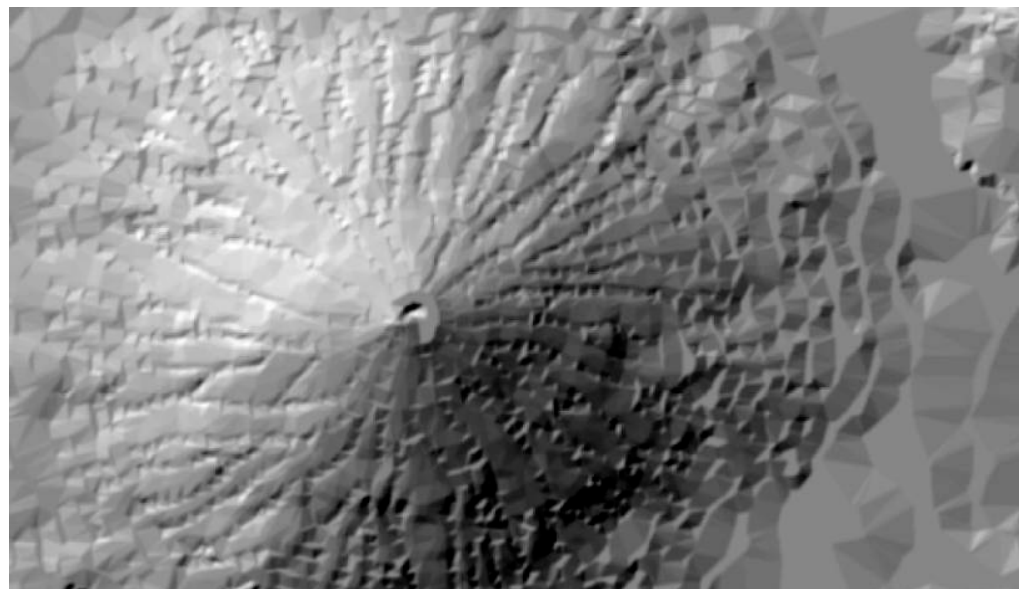
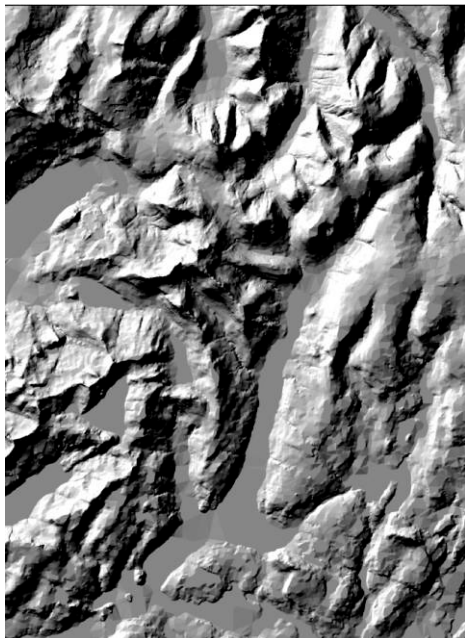
## - Samantekt -

V201111055

**Anna Guðrún Ahlbrecht**

**Landmælingar Íslands**

**Akranesi 02.01.2012**



## Landupplýsingar og gæði

Ýmsar stofnanir og fyrirtæki hafa í mörg ár unnið að gerð gagnagrunna sem halda utan um ýmsar landupplýsingar. Eftir því sem tækninni fleytir fram hefur þróunin á þessu sviði verið hröð og gagnamagn aukist í samræmi við það. Þá er mikilvægt að huga að gæðum gagnanna og í hvaða tilgangi þau voru búin til.

Landupplýsingar eru notaðar í ýmsum tilgangi og stundum ekki í þeim tilgangi sem tilteknir gagnagrunnar voru hugsaðir fyrir í upphafi. Í þessu samhengi getur t.d. aldur gagnanna verið mikilvægur, nákvæmni þeirra eða mælikvarði. Stundum henta gögnin ekki í því samhengi sem þau eru notuð. Þetta er ekki alltaf augljóst og því mikilvægt að upplýsingar um gögnin (landupplýsingarnar) fylgi, séu tiltæk og auðlæsileg.

Tækninefndin ISO TC 211 (Geographic information/Geomatics) hefur þróað staðla sem eru leiðbeinandi fyrir þá sem vilja lýsum gæðum gagnagrunna sem halda utan um landupplýsingar. Eftir því sem fleiri tileinka sér þessa aðferðafræði því æfðari verða notendur og framleiðendur landupplýsinga að gera sér grein fyrir skilgreindum gæðum gagnagrunna sem verið er að vinna með.

### Skráning um gæði – til hvers þegar hægt er að spyrjast fyrir?

Í flóknara umhverfi en við hér á Íslandi búum við er auðvelt að sjá fyrir sér hversu mikilvægt það er að fylgja þessum samræmdu kröfum sem er lýst í gæðastöðlunum. Hér á Íslandi er þetta samfélag einfaldara og saga gagnagrunnanna ekki löng. Enn man fólk hver bjó til ákveðin gögn, hvað er á bak við þau og í hvaða tilgangi þau voru gerð. En þetta breytist fljótt, sérstaklega þar sem um er að ræða rafræn gögn og afleidd gögn. Ekki er hægt að lesa handbragðið (eins og á teiknuðum kortum) og skriflegar upplýsingar um gögnin eru af skorum skammti. Því fyrr sem farið er að skrá upplýsingar um gögnin, því betra.

### Staðlar um lýsigögn og gæði landupplýsinga

Staðlaröðin ISO 19100 fjallar um landupplýsingar og eftirtaldir staðlar úr röðinni snerta gæðamál og lýsigögn:

ISO 19113	Quality principles / Gæðareglur
ISO 19114	Quality evaluation procedures / Gæðamatsaðferðir
ISO 19115	Metadata / Lýsigögn
ISO 19115-2	Metadata – Part 2: Extensions for imagery and gridded data / Lýsigögn – 2. hluti: Viðauki fyrir mynd- og rastagögn.
ISO 19131	Data product specifications / Gagnaafurðalýsing
ISO 19138	Data quality measures / Gagnagæðamat
ISO 19139	Metadata – Implementation specification / Lýsigögn útfærslulýsing

Nú er verið að búa til staðal sem sameinar staðlana 19113, 19114 og 19138 í þeim tilgangi að einfalda og samræma kröfur og bæta aðgengið að gæðastöðlunum. Nýi staðallinn er númer 19157 en er enn á frumvarpsstigi sem þýðir að enn getur innihald hans breyst. Skjalið er sem sagt ekki orðið að viðurkenndum staðli.

Megininntak í gæðastöðlun fyrir landupplýsingar er að framleiðendur og notendur „tali sama tungumálið“, þ.e. að ákveðin skilgreind hugtök séu notuð svo ekki fari á milli mála um hvað er verið að fjalla. Staðallinn ISO 19113 var einmitt búinn til í þessum tilgangi. Gæðunum er lýst með gæðastökum (e. data quality element) og yfirlitsstökum gagnagæða (e. data quality overview elements). ISO 19114 er notaður í samhengi við öflun gagna og geymslu gagnasafna og til að meta gögn. Útkoman er síðan skráð samkvæmt 19115 í lýsigögn eða í sérstaka gæðaskýrslu. Kröfurnar sem eru gerðar í 19115 miðast við að auðvelt sé að finna og nýta gagnasett. Markmiðið með 19138 Gagnagæðamat er að leiðbeina framleiðandanum við val á gæðamatsaðferð fyrir skráningu á gæðum gagnanna. Um leið er verið að auðvelda notandanum mat á notagildi gagnasetta fyrir þá. 19131 nýtist sem eins konar gátlisti því í honum er upptalning atriða sem ættu að koma fram í gagnalýsingu.

## Nánar um hvern og einn staðal

### ISO 19113 Quality principles / Gæðareglur

Venjulega eru gagnasett (e. datasets) ekki framleidd í einum ákveðnum tilgangi en gert ráð fyrir að hægt sé að nota þau í ýmsum tilgang. Staðallinn er m.a. hugsaður fyrir framleiðendur landupplýsinga til að hjálpa til við gerð afurðalýsinga (e. product specification).

Til þess að auðvelda mat á gagnasettum er nauðsynlegt að þekkja gæðastökin (e. quality elements) og yfirlitsstök gagnagæðanna (e. data quality overview element). Tilgangur staðalsins er ekki að lýsa hvernig eigi að mæla muninn á gagnasettinu og alheiminum (raunveruleikanum). Hann skilgreinir flokkunarfræði fyrir mismun sem er venjulega eru framkvæmdar mælingar á. Lýsiorðin sem notuð eru yfir mismuninn eru talin upp í gæðastökunum (quality elements) og *undirþáttum gagnagæða* (data quality subelement) sem heyrja þar undir.

Staðallinn er aðeins leiðbeinandi og segir ekki til um lágmarks gæðakröfur. Þó bendir staðallinn á að ef að vísa eigi til þess að ákveðin framleiðsla (gögn) uppfylli kröfur ISO 19113, verði að uppfylla ákveðnar kröfur sem settar eru fram í viðauka A í staðlinum. Þar er ýmsum prófunum er lýst.

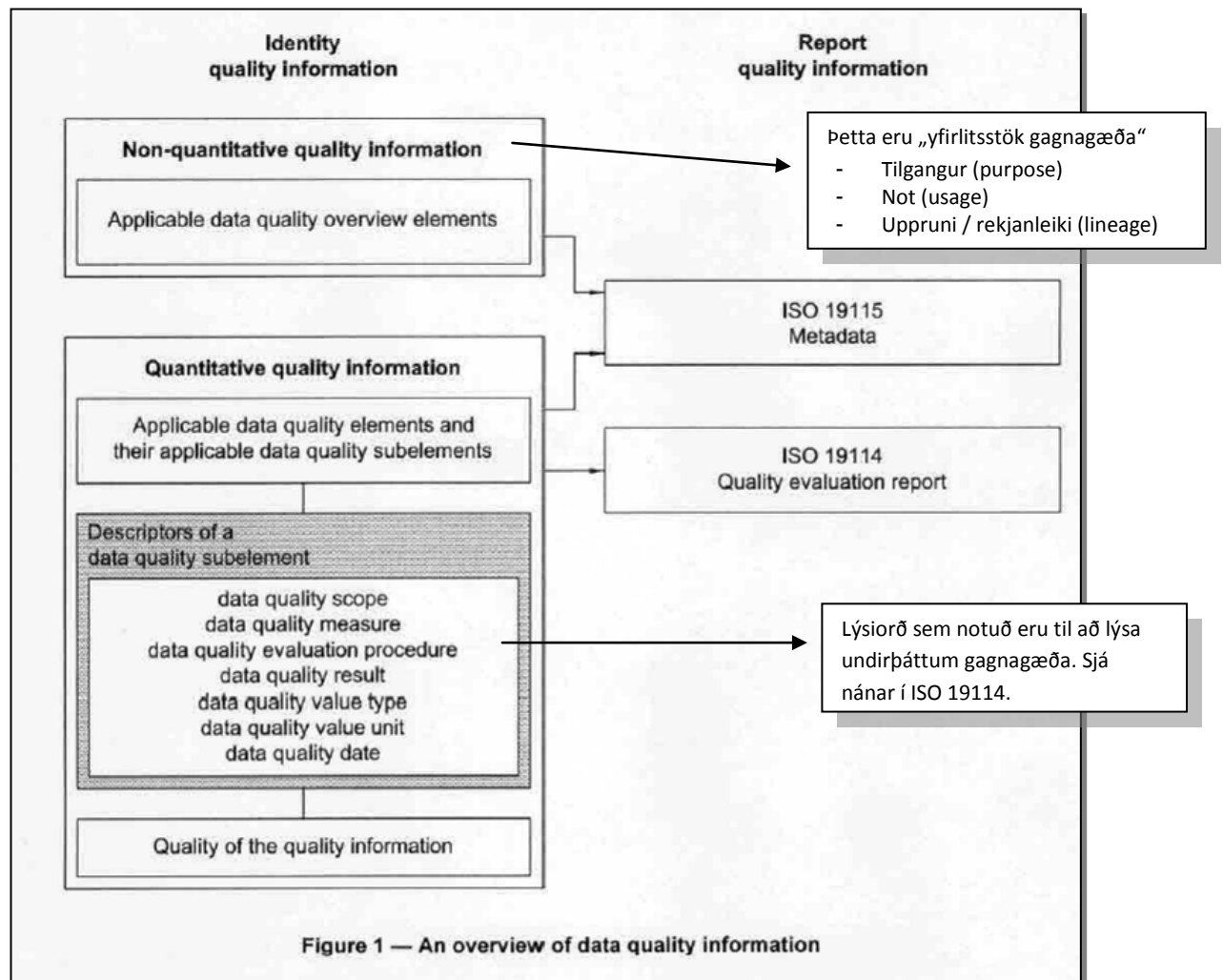
Þó svo að staðallinn sé í grunninn hugsaður fyrir stafræn landfræðileg gögn, er hægt að yfirfæra grunnhugmynd staðalsins á aðrar tegundir landfræðilegra gagna, s.s. útgefin kort og textaskjöl.

### Gagnagæðastök

Í staðlinum eru leiðbeiningar fyrir lýsingar á gæðum gagnasafna. Farið er yfir þá þætti sem er nauðsynlegt að skrá til að lýsa gæðum gagna.

Staðallinn segir að gæðum gagna skuli lýst með eftirfarandi tveimur þáttum:

- *Gagnagæðastök* (data quality element) og *undirþáttur gagnagæða* (data quality subelement) sem heyrja þar undir
- *Yfirlitsstök gagnagæða* (data quality overview element).



Mynd 1. Myndin er úr staðlinum ISO 19113 og sýnir tengingar gæðastaðlanna.

Í *yfirlitsstököm gagnagæða* (data quality overview elements) eru skilgreind þrjú stök sem innihalda almennar lýsingar. Þetta eru gæðaupplýsingar sem ekki hafa mælanleg gildi (non quantitative):

- *Tilgangur* (purpose): í hvaða tilgangi var gagnasettið búið til og upplýsingar skráðar um fyrirhugaða notkun þess.
- *Not* (usage): lýst er til hvaða nota gagnasett hefur verið (hjá framleiðanda og kaupanda / notanda).
- *Uppruni / rekjanleiki* (lineage): lýst er sögu gagnasetts, frá öflun og allt um notkunarferlið og breytingaferli (breytt notkun frá uppruna o.fl.).

*Gagnagæðastökin* (data quality element) eru fimm og eru mælanleg (quantitative). Þeim er er skipt niður í 15 *undirþætti gagnagæða* (data quality subelements) og þau notuð þar sem það á við.

Undirþættirnir lýsa ákveðnum hluta viðkomandi gagnagæðastaks. Fyrir hvert gæðastak eru 2-4 undirþættir taldir upp. Gæðahugtökin (undirþættirnir) sem eru notuð til lýsingar eru aðeins dæmi en ekki endanlegur listi yfir þau. Hægt er að bæta við listann:

	<i>Gagnagæðastök</i>	<i>Skýring</i>	<i>Undirþættir gagnagæðastaka</i>	<i>Skýring</i>
a	<b>„Fullkomleiki“</b>	<b>Completeness</b> - presence and absence of features, their attributes and relationships		
1		umframskráning	<b>commission</b>	excess data present in a dataset
2		vanskráning	<b>omission</b>	data absent from a dataset
b	<b>Gæðastig gagna</b>	<b>Logical consistency</b> - degree of adherence to logical rules of data structure, attribution and relationships (data structure can be conceptual, logical or physical)		
3		fylgni við hugtakakerfið	<b>conceptual consistency</b>	adherence to rules of the conceptual schema
4		fylgni við gildisóðal	<b>domain consistency</b>	adherence of values to the value domains
5		formfesta	<b>format consistency</b>	degree to which data is stored in accordance with the physical structure of the dataset.
6		grenndar samkvæmni	<b>topological consistency</b>	correctness of the explicitly encoded topological characteristics of a dataset.
c	<b>Staðsetningarnákvæmni</b>	<b>Positional accuracy</b> - accuracy of the position of features		
7		raunnákvæmni - ytri	<b>absolute or external accuracy</b>	closeness of reported coordinate values to values accepted as or being true
8		afstæð nákvæmni	<b>relative or internal accuracy</b>	closeness of the relative positions of features in a dataset to their respective relative positions accepted as or being true
9		Staðsetningarnákvæmni rasteragagna	<b>gridded data position accuracy</b>	closeness of gridded data position values to values accepted as or being true.
d	<b>Tímanákvæmni gagna</b>	<b>Temporal accuracy</b> accuracy of the temporal attributes and temporal relationships of features		
10		nákvæmni tímamælingar	<b>accuracy of a time measurement</b>	correctness of the temporal references of an item (reporting of error in time measurement)
11		tímasamkvæmni	<b>temporal consistency</b>	correctness of ordered events or sequences, if reported
12		réttmæti tíma	<b>temporal validity</b>	validity of data with respect to time.
e	<b>Þemanákvæmni gagna</b>	<b>(thematic accuracy)</b> accuracy of quantitative attributes and the correctness of non-quantitative attributes and of the classifications of features and their relationships		
13		réttleiki flokkunar	<b>classification correctness</b>	comparison of the classes assigned to features or their attributes to a universe of discourse (e.g. ground truth or reference dataset)
14		réttleiki eiginda sem ekki eru mælanlegar	<b>non-quantitative attribute correctness</b>	correctness of non-quantitative attributes
15		réttleiki mælanlegra eiginda	<b>quantitative attribute accuracy</b>	accuracy of quantitative attributes.

### Hvernig undirbáttum gagnagæðastaka skal lýst

Upplýsingarnar um undirbætti gagnagæðastaka (data quality subelements) eru skráðar og til lýsingar eru eftirfarandi lýsiorð notuð (sbr. ISO 19114):

- data quality scope
- data quality measure
- data quality evaluation procedure
- data quality result
- data quality value type
- data quality value unit
- data quality date

Í staðlinum er stöðugt vakin athygli á mælanlegum (quantitative) og ekki mælanlegum (non-quantitative) upplýsingum.

Staðallinn er nátengdur við ISO19114 og ISO19115, en þessir staðlar lýsa skráum fyrir upplýsingagjöf um gæði. ISO 19138 tengist einnig þessum staðli.

Þessi staðall er einkum ætlaður þeim sem þróa hugbúnað fyrir landupplýsingakerfi og framleiðendum/dreifiaðilum landupplýsinga.

#### ISO 19114 Quality evaluation procedures / Gæðamatsaðferðir

Þessi staðall lýsir því hvernig skrá á gæði þar sem ólík gæðatriði eru skráð með ólíkum tölugildum. Staðallinn er rammi fyrir gæðamatsferli fyrir gagnasöfn landfræðilegra gagna. Þannig geta gagnaframleiðendur skilgreint hversu vel framleiðsla þeirra fellur að framleiðslulýsingu staðsins. Notendur geta einnig skilgreint þarfir sínar og hversu vel þeim er mætt. Staðallinn er þannig viðmiðunarrammi fyrir ákvörðun og mat á gæðum á gagnasafni (dataset) í samræmi við grunnreglur ISO 19113.

Staðallinn lýsir ferli í slíku mati og býður uppá nokkrar aðferðir til þessa. Auk þess er lýst hvernig gæði eru skráð, ýmist í matsskýrslu eða með lýsigögnum.

Þessi staðall er einkum ætlaður þeim sem þróa hugbúnað fyrir landupplýsingakerfi og framleiðendum/dreifiaðilum landupplýsinga.

#### ISO 19115 Metadata / Lýsigögn

Lýsigögn eru til að aðstoða notendur við að finna gögn sem henta í tiltekin verkefni. Í þessum lýsigagnastaðli er talið upp hvaða lýsigögn eigi að fylgja með gögnum. Ákveðin stök eru skilgreind og nokkur þeirra eru tilgreind sem skylduskráningar atriði (e.mandatory - M). Önnur eru valkvæm (e.optional - O) og enn önnur er skylda að skrá í vissum tilvikum eða undir vissum kringumstæðum (e.conditional - C).

Table 3 — Core metadata for geographic datasets

<b>Dataset title (M)</b> (MD_Metadata > MD_DataIdentification.citation > CI_Citation.title)	<b>Spatial representation type (O)</b> (MD_Metadata > MD_DataIdentification.spatialRepresentationType)
<b>Dataset reference date (M)</b> (MD_Metadata > MD_DataIdentification.citation > CI_Citation.date)	<b>Reference system (O)</b> (MD_Metadata > MD_ReferenceSystem)
<b>Dataset responsible party (O)</b> (MD_Metadata > MD_DataIdentification.pointOfContact > CI_ResponsibleParty)	<b>Lineage (O)</b> (MD_Metadata > DQ_DataQuality.lineage > LI_Lineage)
<b>Geographic location of the dataset (by four coordinates or by geographic identifier) (C)</b> (MD_Metadata > MD_DataIdentification.extent > EX_Extent > EX_GeographicExtent > EX_GeographicBoundingBox or EX_GeographicDescription)	<b>On-line resource (O)</b> (MD_Metadata > MD_Distribution > MD_DigitalTransferOption.onLine > CI_OnlineResource)
<b>Dataset language (M)</b> (MD_Metadata > MD_DataIdentification.language)	<b>Metadata file identifier (O)</b> (MD_Metadata.fileIdentifier)
<b>Dataset character set (C)</b> (MD_Metadata > MD_DataIdentification.characterSet)	<b>Metadata standard name (O)</b> (MD_Metadata.metadataStandardName)
<b>Dataset topic category (M)</b> (MD_Metadata > MD_DataIdentification.topicCategory)	<b>Metadata standard version (O)</b> (MD_Metadata.metadataStandardVersion)
<b>Spatial resolution of the dataset (O)</b> (MD_Metadata > MD_DataIdentification.spatialResolution > MD_Resolution.equivalentScale or MD_Resolution.distance)	<b>Metadata language (C)</b> (MD_Metadata.language)
<b>Abstract describing the dataset (M)</b> (MD_Metadata > MD_DataIdentification.abstract)	<b>Metadata character set (C)</b> (MD_Metadata.characterSet)
<b>Distribution format (O)</b> (MD_Metadata > MD_Distribution > MD_Format.name and MD_Format.version)	<b>Metadata point of contact (M)</b> (MD_Metadata.contact > CI_ResponsibleParty)
<b>Additional extent information for the dataset (vertical and temporal) (O)</b> (MD_Metadata > MD_DataIdentification.extent > EX_Extent > EX_TemporalExtent or EX_VerticalExtent)	<b>Metadata date stamp (M)</b> (MD_Metadata.dateStamp)

Mynd 2. Taflan er úr staðlinum ISO 19115 staðlinum og sýnir grunnatriðin sem ætti að skrá í lýsigögn (e. core metadata).

Í viðauka eru skilgreind yfir 400 stök og einnig er skilgreint hvernig hægt er að útvíkka staðalinn með nýjum stökum. Þessi staðall er einkum ætlaður framleiðendum og dreifingaraðilum landupplýsinga sem og þeim sem þróa hugbúnað fyrir landupplýsingakerfi.

## ISO 19115-2 Metadata Part 2: Extensions for imagery and gridded data

### Lýsigögn 2. hluti: Viðauki fyrir mynd- og rastagögn

Þessi staðall eykur að mörgu leiti við ISO 19115 hvað varðar myndir og aðrar gerðir af rastagögnum. Skilgreind eru lýsigagnastök sem styðja mynd- og rastagögn og að auki er bætt við skilgreiningar UML líkansins frá ISO 19115 staðlinum.

Staðallinn er einkum ætlaður þeim sem þróa hugbúnað fyrir landupplýsingakerfi og framleiðendum/dreifiaðilum landupplýsinga.

### ISO 19131 Data product specifications / Gagnaafurðalýsing

Þessi staðall lýsir fjölskyldu gagnalýsingastaðla. Teknir eru saman þættir frá öðrum stöðlum og safnað saman í þennan staðal. Staðallinn skilgreinir gagnaframleiðslu og innihald hennar, hann lýsir kröfum sem eru gerðar til skráningar landupplýsinga. Þessum þáttum er lýst af mikilli nákvæmni í staðlinum:

- Skylda:
- a) Yfirlit (overview)
  - b) Sérhæft svið (specification scope)
  - c) Auðkenni gagnaframleiðslu (Data product identification)
  - d) Innihald gagna og uppbygging (Data content and structure)
  - e) Viðmiðunarkerfi (Reference system)
  - f) Gæði gagna (Data quality)
  - g) Afhending gagnaframleiðslu (Data product delivery)
  - h) Lýsigögn (Metadata)

- Valkvæm stök:
- i) Gagnataka (Data capture)
  - j) Viðhald gagna (Data maintenance)
  - k) Birting (Portrayal)
  - l) Aðrar upplýsingar (Additional information)

### ISO 19138 Data quality measures / Gagnagæðamat

Þessi staðall er hluti af gæðastaðlaröðinni 19113, 19114 og þeim hluta sem varðar gæðaupplýsingar eins og þær eru settar fram í 19115. Hann byggir einnig á 19135 og þá umfram allt hvað varðar skráningu á gæðapáttum.

Þetta er tæknileg lýsing / greining (TS) á líkani fyrir gæðapætti með 13 stökum og skilgreinir auk þess röð undirgæðapátta.

Þarna kemur fram skilgreining á röð gæðaatríða sem hægt er að nota til að setja tölugildi á gæðastök sem eru skilgreind í 19113. Aðferðafræðinni við gæðamatið er lýst í 19114. Þetta er ekki fullkominn listi yfir öll atríði heldur byrjun á slíkum vinnulistum. Lýsingarnar þurfa að vera mjög nákvæmar svo að ólíkir aðilar sem framkvæma gæðaprófun geti komist að sömu niðurstöðu um gæði gagnanna, óháðir hvorum öðrum.

Þessi staðall / TS er einkum ætlaður þeim sem þróa landupplýsingavörur, þeim sem þróa hugbúnað fyrir landupplýsingakerfi, framleiðendum/dreifiaðilum landupplýsinga, notendum landupplýsinga og LUK og þeim sem þróa staðla.

### ISO 19139 Metadata – implementation specification / Lýsigögn - útfærslulýsing

Þessi hluti af 19100 stöðlunum hefur að geyma kóðun fyrir XML (spatial metadata eXtensible Markup Language (smXML)) landfræðileg lýsigögn. Innleiðinguna má rekja til ISO 19115. Tilgangurinn með staðlinum er að stuðla að aukinni samhæfingu og er viðbót við ISO 19115. Með staðlinum er þörfin fyrir kóðun á lýsigögnum og samskipti milli ólíkra lýsigagnaskráa o.þ.h. uppfyllt. Þessi tæknilega útlistun (TS) byggir einnig á grunnhugmyndinni um líkanakeyrða uppbyggingu.





Lýsigögnin innihalda upplýsingar um auðkenni, skilyrði, umfang, gæði, rúmfræðileg og tímaleg viðmið, dreifingu, uppruna (lineage) og viðhald rafræns landfræðilegs gagnasafns. ISO TS 19139 er hannað til að leggja til almenna XML gerð fyrir lýsingu, mat og samskipti með landfræðileg lýsigögn.

Þessi staðall er einkum ætlaður þeim sem þróa hugbúnað fyrir landupplýsingakerfi og framleiðendum/dreifiaðilum landupplýsinga.