YYT-C3001 Management of environmental data and information

Lecture 7: Spatial Metadata and data quality

Aalto University
School of Engineering

Jussi Nikander

7.10.2020

Contents of this lecture

Spatial metadata

- **The ISO 19115 standard**
- **INSPIRE metadata**

Spatial data quality issues

Data interoperability

Learning goals for this lecture

Understand the basic tenets of ISO 19115 metadata standard and the Catalog service for the web interface

Understand what is spatial data quality and be able to name and describe basic data quality characteristics

Know how interoperability can be characterized in different ways

Spatial metadata

What Metadata?

- Spatial data typically has a complex structure
- This structure needs to be known (typically in a machine-readable format) in order to exploit spatial data
- In addition, the users need to know what a given spatial dataset represents, and what it therefore can be used for
- Therefore metadata (data about the data) is important for spatial datasets

Aalto University
School of Engineering


```
<gmd:MD_Metadata xsi:schemaLocation="http://www.isotc211.org/2005/gmd ../../../../web/geonetwork/xml/schemas/iso19139/schema.xsd">
  <gmd:fileIdentifier>
    <gco:CharacterString>a959b07c-16ff-4ed0-9a3b-72d6bbc026ff</gco:CharacterString>
  </gmd:fileIdentifier>
  <gmd:language>
    <gco:CharacterString>fin</gco:CharacterString>
  </gmd:language>
  <gmd:hierarchyLevel>
    <gmd:MD_ScopeCode codeListValue="series" codeList="http://standards.iso.org/iso/19139/resources/gmxCodelists.xml#MD_ScopeCode"/>
  </gmd:hierarchyLevel>
  <gmd:hierarchyLevelName>
    <gco:CharacterString>Aineisto</gco:CharacterString>
  </gmd:hierarchyLevelName>
  <gmd:contact>
    <gmd:CI_ResponsibleParty>
      <gmd:organisationName xsi:type="gmd:PT_FreeText_PropertyType">
        <gco:CharacterString>Maanmittauslaitos</gco:CharacterString>
        <gmd:PT_FreeText>
          <gmd:textGroup>
            <gmd:LocalisedCharacterString locale="#FI">Maanmittauslaitos</gmd:LocalisedCharacterString>
          </gmd:textGroup>
          <gmd:textGroup>
            <gmd:LocalisedCharacterString locale="#EN">National Land Survey of Finland</gmd:LocalisedCharacterString>
          </gmd:textGroup>
          <gmd:textGroup>
            <gmd:LocalisedCharacterString locale="#SV">Lantmäteriverket</gmd:LocalisedCharacterString>
          </gmd:textGroup>
        </gmd:PT_FreeText>
      </gmd:organisationName>
      <gmd:contactInfo>
        <gmd:CI_Contact>
          <gmd:address>
            <gmd:CI_Address>
              <gmd:electronicMailAddress xsi:type="gmd:PT_FreeText_PropertyType">
                <gco:CharacterString>asiakaspalvelu@maanmittauslaitos.fi</gco:CharacterString>
                <gmd:PT_FreeText>
                  <gmd:textGroup>
                    <gmd:LocalisedCharacterString locale="#FI">asiakaspalvelu@maanmittauslaitos.fi</gmd:LocalisedCharacterString>
                  </gmd:textGroup>
                  <gmd:textGroup>
                    <gmd:LocalisedCharacterString locale="#EN">customerservice@nls.fi</gmd:LocalisedCharacterString>
                  </gmd:textGroup>
                  <gmd:textGroup>
                    <gmd:LocalisedCharacterString locale="#SV">kundservice@lantmateriverket.fi</gmd:LocalisedCharacterString>
                  </gmd:textGroup>
                </gmd:PT_FreeText>
              </gmd:electronicMailAddress>
            </gmd:CI_Address>
          </gmd:address>
        </gmd:CI_Contact>
      </gmd:contactInfo>
      <gmd:role>
        <gmd:CI_RoleCode codeList="http://standards.iso.org/iso/19139/resources/gmxCodelists.xml#CI_RoleCode" codeListValue="owner"/>
      </gmd:role>
    </gmd:CI_ResponsibleParty>
  </gmd:contact>
  <gmd:dateStamp>
    <gco:DateTime>2018-09-03T15:00:33</gco:DateTime>
  </gmd:dateStamp>
  <gmd:locale>
    <gmd:PT_Locale id="SV">
      <gmd:languageCode>
        <gmd:LanguageCode codeList="http://www.loc.gov/standards/iso639-2/" codeListValue="swe"/>
      </gmd:languageCode>
      <gmd:characterEncoding>
        <gmd:MD_CharacterSetCode codeList="http://standards.iso.org/iso/19139/resources/gmxCodelists.xml#MD_CharacterSetCode" codeListValue="UTF-8"
      </gmd:characterEncoding>
    </gmd:PT_Locale>
  </gmd:locale>
```

Metadata for two different purposes

- **Machine-readable metadata promotes machine-to-machine interoperability of spatial data**
- **Client can search for specific kinds of datasets automatically**
- **Software knows what the data set represents and how to visualize it**

- **Human-readable metadata allows the data set user to know what they're working with**
- **What, exactly the data in the set represents**
  - What is the meaning (semantics) of the data values

Machine-readable metadata

- **Machine-readable metadata requires a metadata standard**
  - This allows the data to be described in a format that can be interpreted
- **Metadata standards make it possible to automatically combine data from different sources and make more complex services**

Human-readable metadata

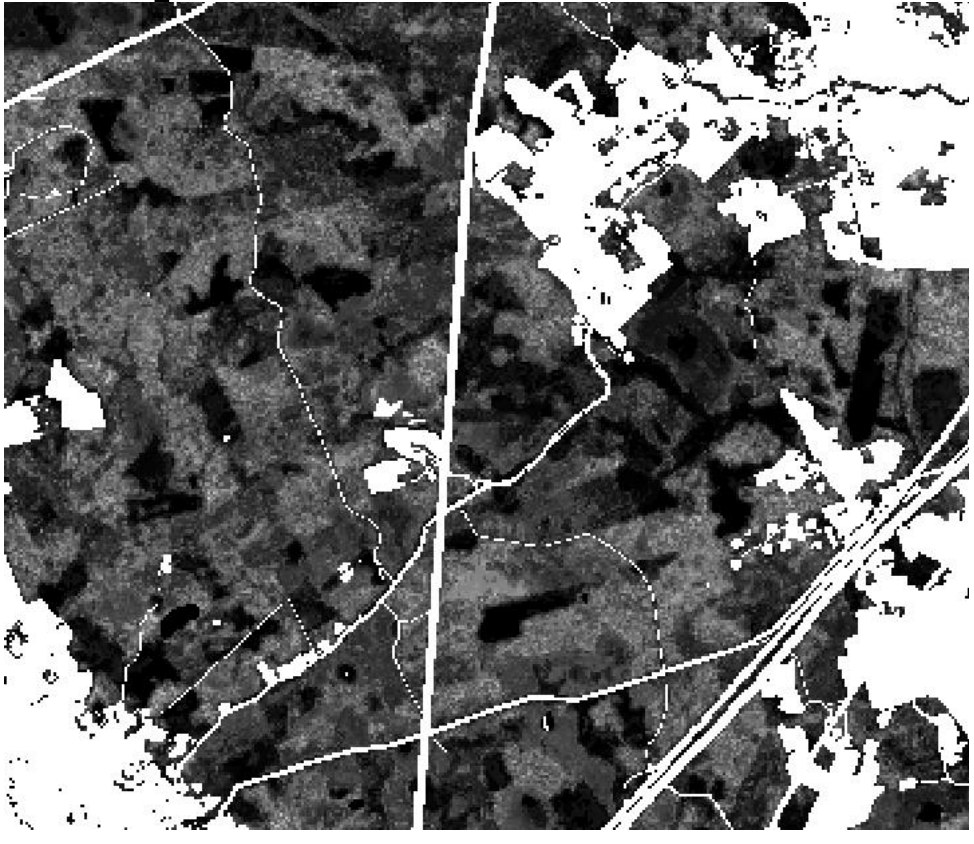
- **Human-readable metadata typically contains descriptions of the data in natural language**
- **The description can be in a specified form, or free-form**
  - Using a specific form typically makes it easier for the reader
  - The form can be specified in a metadata standard

- **Simple data example: what does 298.25 mean?**
- **If you know it is a measure of temperature, does that help you know what it means?**
- **What about if you know it is a measure of temperature in the Kelvin scale?**
- **298.25 K = 25.1 C**

Human-readable metadata: a spatial data example

“The estimation errors at pixel level are rather high but decrease when the area in question increases, i.e., when the area of interest consists of several pixels.”

- **Using this data for analysis on individual pixel level may be a bad idea**
- **The (potentially) large errors on single pixels will persist through the analysis**
- **When aggregated into larger areas in the analysis, the errors even out**



Data: Multi-source national forest inventory 2017, total volume of trees

ISO 19115: Geographic Information – Metadata

- **ISO Standard for how to represent spatial metadata**
- **Contains definitions of**
  - metadata elements
  - Extensions for imagery and gridded data
  - XML schema
- **Is used as a basis for other metadata standard definitions**
  - E.g. Finnish JHS 158
- **Original standard is from 2003, the current revised version was finalized in 2014**

ISO 19115 contents

- **The elements of the 2014 version of the standard consist of**
  - Metadata information
  - Identification information
  - Constraint information
  - Lineage information
  - Maintenance information
  - Spatial representation information
  - Reference system information
  - Content information
  - Portrayal catalogue information
  - Distribution information
  - Metadata extension information
  - Application schema information
  - Service metadata information

ISO 19115 in practice

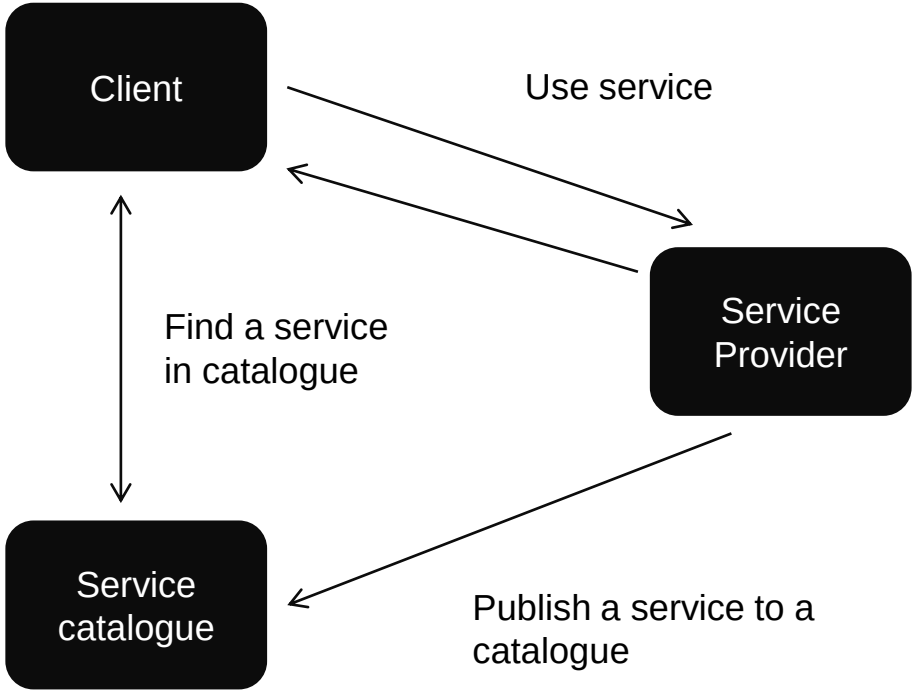
- **The standard defines an XML schema that should be used when using the standard**
  - If you want a machine-readable implementation
- **Many of the fields in the standard contain natural language text**
  - E.g. constraint information contains a text string of free-form text
- **In a specific application, relevant parts of the standard should be used**

Aalto University
School of Engineering

```
<mdb:metadataIdentifier/>
<mdb:contact gco:nilReason="inapplicable"/>
<mdb:dateInfo gco:nilReason="inapplicable"/>
<mdb:identificationInfo>
 <mri:MD_DataIdentification>
  <mri:citation>
   <cit:CI_Citation>
    <cit:title/>
   </cit:CI_Citation>
  </mri:citation>
  <mri:abstract/>
  <!-- example of Use Constraint -->
  <mri:resourceConstraints>
   <mco:MD_Constraints>
    <mco:useLimitation>
     <gco:CharacterString>Description of use limitation</gco:CharacterString>
    </mco:useLimitation>
    <mco:useLimitation></mco:useLimitation>
    <mco:constraintApplicationScope>
     <mcc:MD_Scope>
      <mcc:level>
       <mcc:MD_ScopeCode codeList="009" codeListValue="feature"></mcc:MD_ScopeCode>
      </mcc:level>
     </mcc:MD_Scope>
    </mco:constraintApplicationScope>
    <mco:graphic>
     <mcc:MD_BrowseGraphic>
      <mcc:fileName>
       <gco:CharacterString>File name of graphic for the constraint</gco:CharacterString>
      </mcc:fileName>
      <mcc:fileDescription>
       <gco:CharacterString>Description of graphic for the constraint</gco:CharacterString>
      </mcc:fileDescription>
      <mcc:fileType>
       <gco:CharacterString>File type for the graphic for the constraint</gco:CharacterString>
      </mcc:fileType>
      <!-- not supported
      <mcc:imageConstraints></mcc:imageConstraints>
      -->
```

Catalog Service for the Web (CSW)

- **A service for publishing and searching metadata, services, and related objects**
  - The metadata can be further queried
  - Metadata usable both automatically and by users
- **Defined by OGC**
  - Part of the WFS/WMS/etc. family of web service interfaces



CSW over HTTP

- **When used over the HTTP protocol, CSW offers the following functionality**
  - getCapabilities
  - getDomain
  - getRecords
  - getRecordById
  - transaction
  - harvest
  - unHarvest

- **Next, let's take a closer look at this functionality in a group work**
- **I'll divide you into breakout rooms – each room has two functions assigned to them**
  - Find out what these functions are used for
- **You have 20 minutes**
- **https://jamboard.google.com/d/17pcFFON-tD-ImiJsk6OOrU93zVFEQvmSXrHBxr5BtC4/edit?usp=sharing**

INSPIRE Metadata

Aalto University
School of Engineering

INSPIRE metadata

- **The INSPIRE directive specifies required metadata requirements for INSPIRE services and data sets**
  - One part of how the SDI promotes interoperability
- **INSPIRE Discovery services need to implement CSW**
  - www.paikkatietohakemisto.fi is the national discovery service in Finland

INSPIRE metadata elements

- **Identification**
- **Classification of spatial data and services**
- **Keywords**
- **Geographic location**
- **Temporal reference**
- **Quality and validity**
- **Conformity**
- **Constraints related to access and use**
- **Organizations responsible for management**
- **Metadata on metadata**

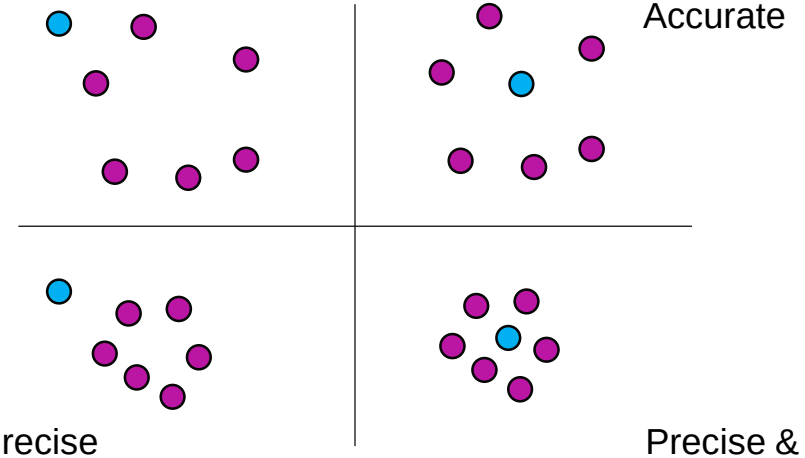
- **The metadata here is based on the ISO 19115 metadata standard**
  - As well as ISO 19119 services
- **You'll take a closer look at this in the exercises**

Spatial Data Quality

Aalto University
School of Engineering

What is spatial data quality

- **Spatial data quality is the degree by which the data satisfies given objectives**

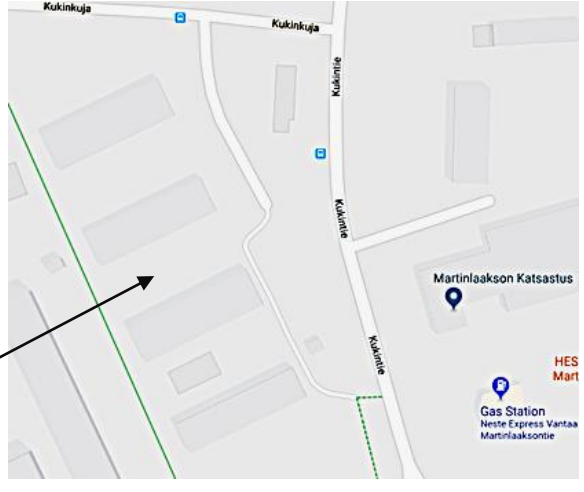


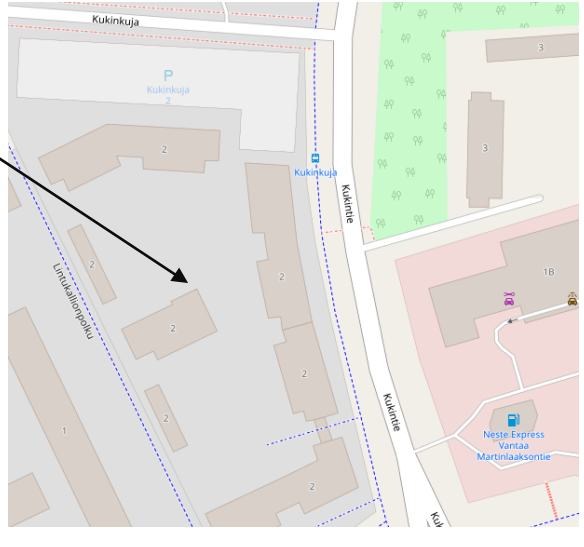
- **Completeness**
  - All required data is there; there is no extraneous data
- **Internal accuracy (precision)**
  - Data elements have similar degree of difference from what they represent
- **External accuracy**
  - Difference between actual values and values in the data set
- **Consistency**
  - There are no conflicts in the data set

ISO 19157: Geographic Information – Data Quality

- **ISO Standard for data quality evaluation**
- **Data quality evaluated based on**
  - Completeness
  - Logical consistency
  - Positional accuracy
  - Thematic accuracy
  - Temporal quality
  - Usability

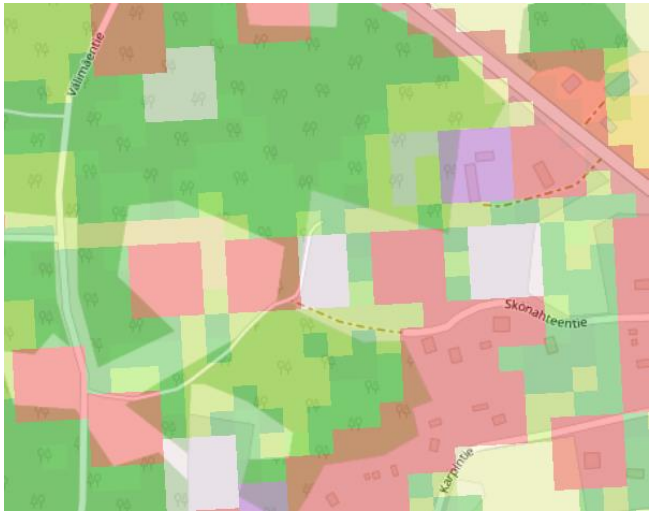
Which one is correct?





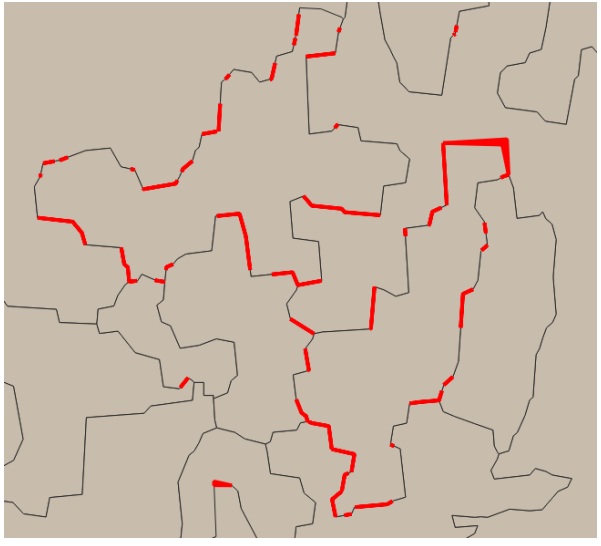
Spatial data quality elements

- **Completeness**
  - Missing data elements are omissions
  - Extranenous data elements are commissions
- **Thematic accuracy**
  - Correctness of attribute data values
- **Temporal quality**
  - Correctness of the time-related attributes
- **Usability**
  - Fitness for purpose



Spatial data quality elements: Logical consistency

- **Topological consistency**
  - Are data elements topologically correct?
- **Conceptual consistency**
  - Do data elements conform to high-level structure of the data set?
- **Domain consistency**
  - Are data values in the accepted range?
- **Format consistency**
  - Are data elements in the format defined by the data set?



Topology problem example: polygons should cover the whole area, but there are small gaps

Spatial data quality group work

- **Now, I'll again divide you into breakout rooms**
- **Each room is given a screenshot of map data**
- **Consider what possible quality problems you can see in the picture**
  - No need to consider quality information not seen in the picture itself

- **You have 10 minutes**
- **https://jamboard.google.com/d/1FumLTF9GWQG-saC_LXS9KD3b-4Gt9kMl888vNFYkHUY/edit?usp=sharing**

Spatial data quality elements: Positional accuracy

Magnitude of the difference is maybe 30-40m (width of the border strip is 10m)

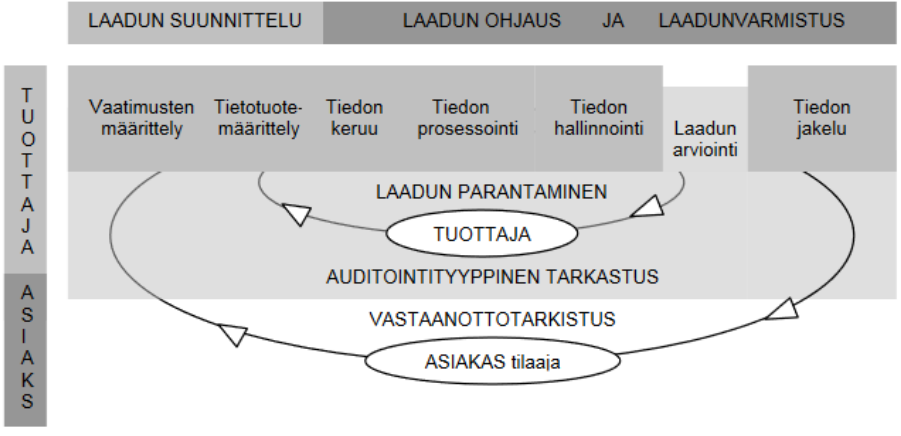
- **External (absolute) accuracy**
  - The difference between data value and the real value
- **Internal (relative) accuracy (precision)**
  - The difference between relative positions of data element values and real elements
- **External accuracy requires reference data**
  - Real world
  - Dataset known to be good

The actual location of the border in the image

The difference does not change (significantly)

JHS 160

- **Finnish spatial data quality standard**
  - Paikkatiedon laadunvarmistus”
- **Based on the old ISO 19113 Spatial Data Quality standard**
- **Describes also the processes of how to create spatial data quality**



Interoperability

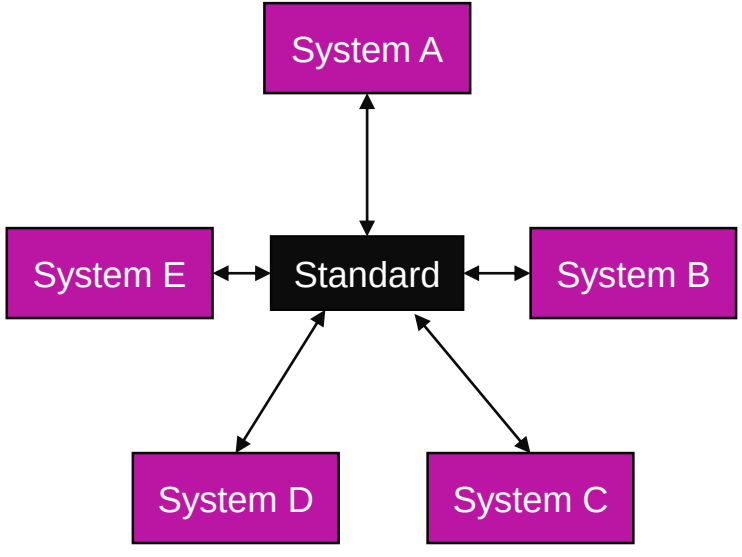
Data transfer, data sharing, and interoperability

When data is transferred from one system to another, the meaning of the data (semantics) and the relationships between different elements in the data (structure) need to be preserved

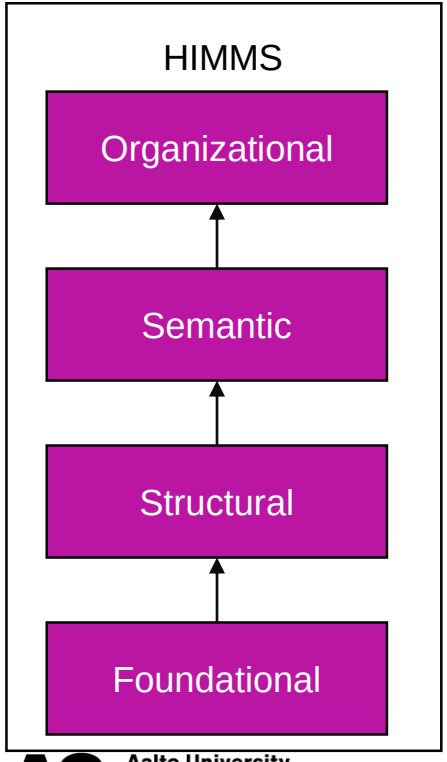
This is a goal of interoperability

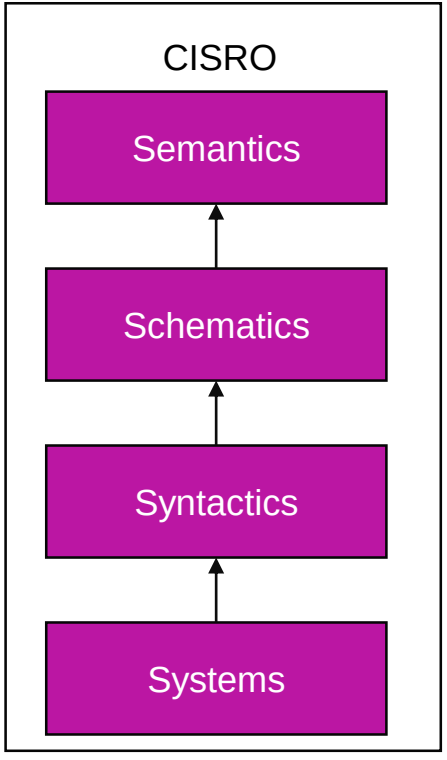
“Data interoperability addresses the ability of systems and services that create, exchange and consume data to have clear, shared expectations for the contents, context and meaning of that data.”
Data interoperability standards consortium

“Interoperability is the ability of different information systems, devices or applications to connect, in a coordinated manner, within and across organizational boundaries to access, exchange and cooperatively use data amongst stakeholders, with the goal of optimizing the health of individuals and populations.”
Healthcare information and management systems society



How interoperability can be defined



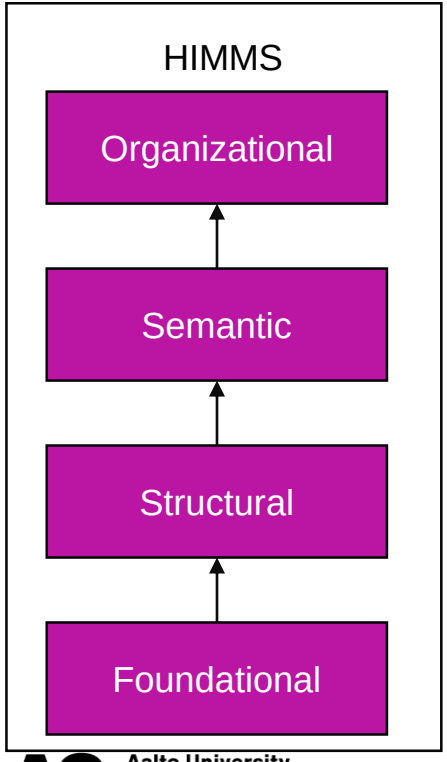


Aalto University
School of Engineering

Source: HIMMS.org: What is interoperability?
Defining the interoperability in the health ecosystem

Source: Box et al. (2015): A Data Specification Framework for the Foundation Spatial Data Framework

How interoperability can be defined

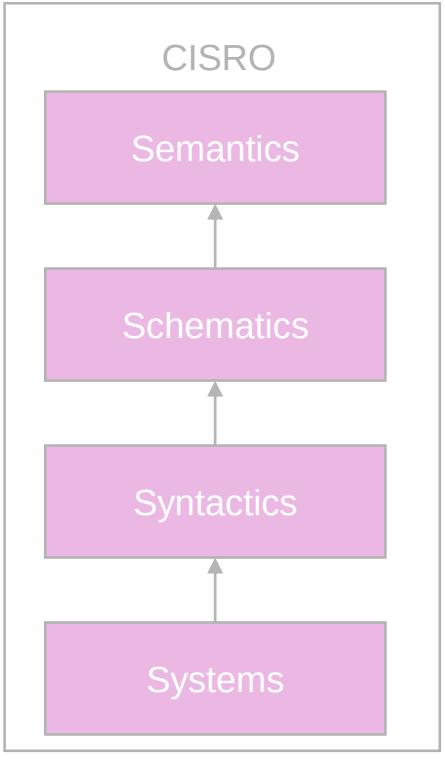


Encompasses the technical components as well as clear policy, social and organizational components

The ability of two or more systems to exchange information and to interpret and use that information

Define the structure or format of data exchange (i.e., the message format standards)

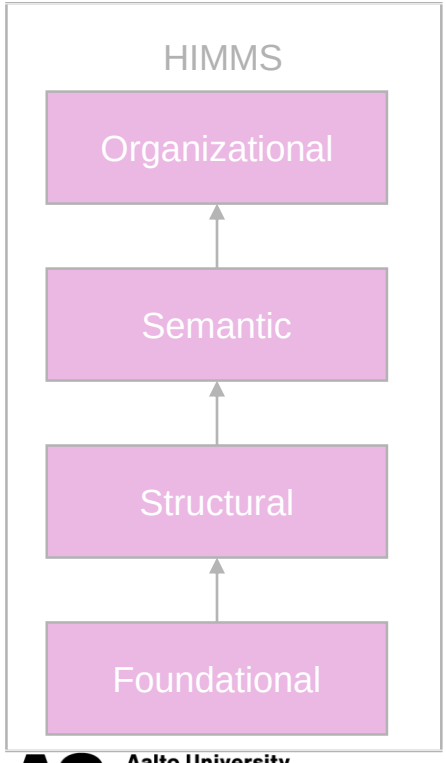
Establish the inter-connectivity requirements needed for one system or application to share data with and receive data from another



Source: HIMMS.org: What is interoperability?
Defining the interoperability in the health ecosystem

Source: Box et al. (2015): A Data Specification Framework for the Foundation Spatial Data Framework

How interoperability can be defined

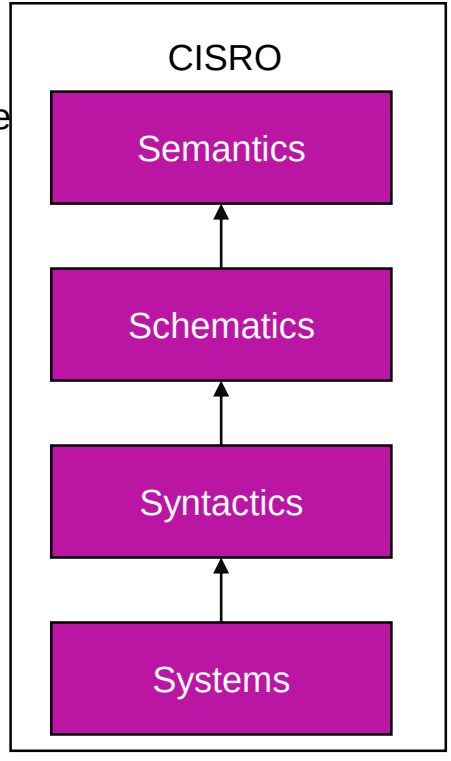


Common meaning for the data (e.g. use of specific vocabularies, ontologies)

Common information exchange formats (e.g. INSPIRE data model, CityGML)

Common data formats and languages (e.g. shapefile, GML, geoJSON)

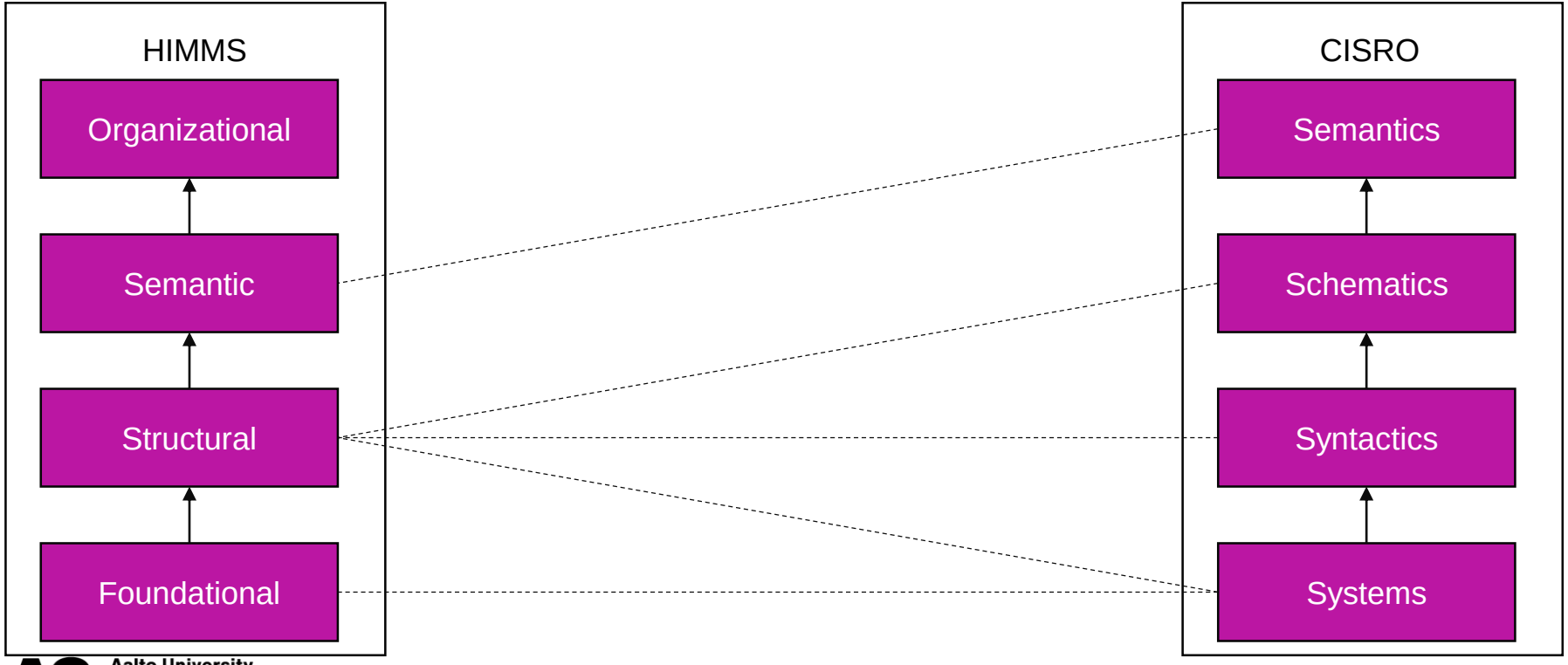
Network protocols (e.g. HTTP), web services (e.g. WDSL), operating systems (e.g. Windows)



Source: HIMMS.org: What is interoperability? Defining the interoperability in the health ecosystem

Source: Box et al. (2015): A Data Specification Framework for the Foundation Spatial Data Framework

How interoperability can be defined

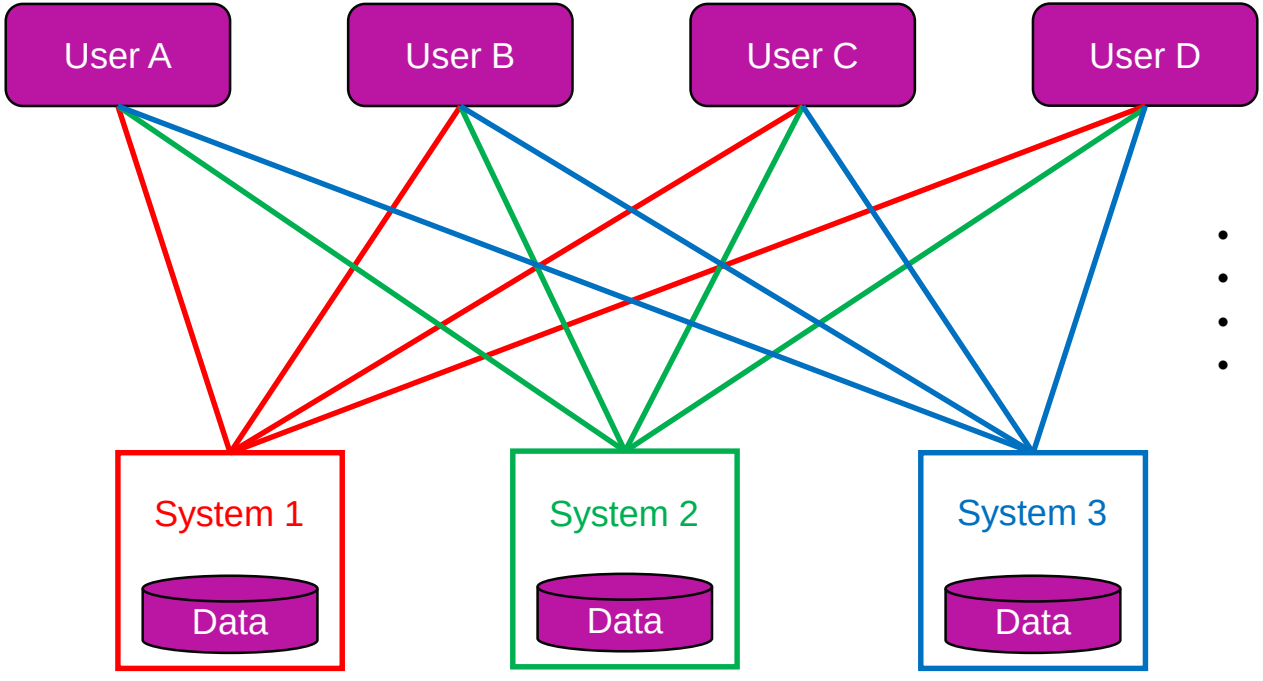


Aalto University
School of Engineering

Source: HIMMS.org: What is interoperability?
Defining the interoperability in the health ecosystem

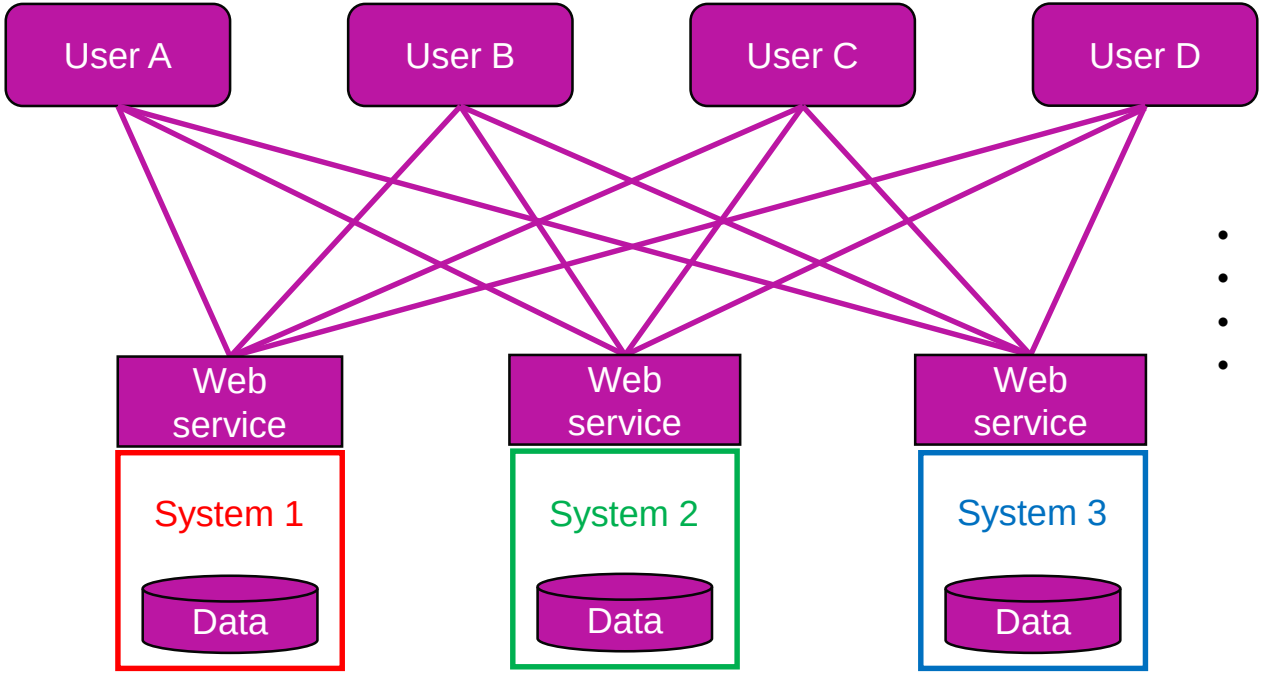
Source: Box et al. (2015): A Data Specification Framework for the Foundation Spatial Data Framework

Different interoperability scenarios



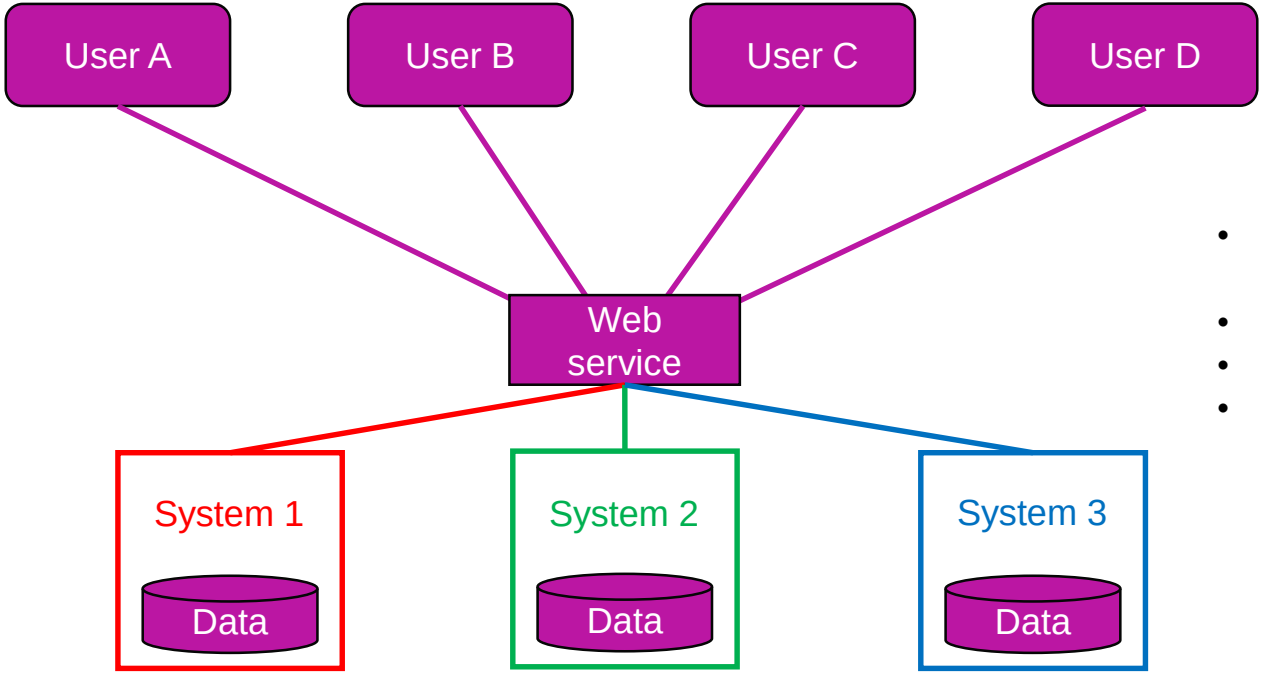
- Different Organizations
- Different semantics
- Different syntactics
- (Foundations and Systems are not that important)

Different interoperability scenarios



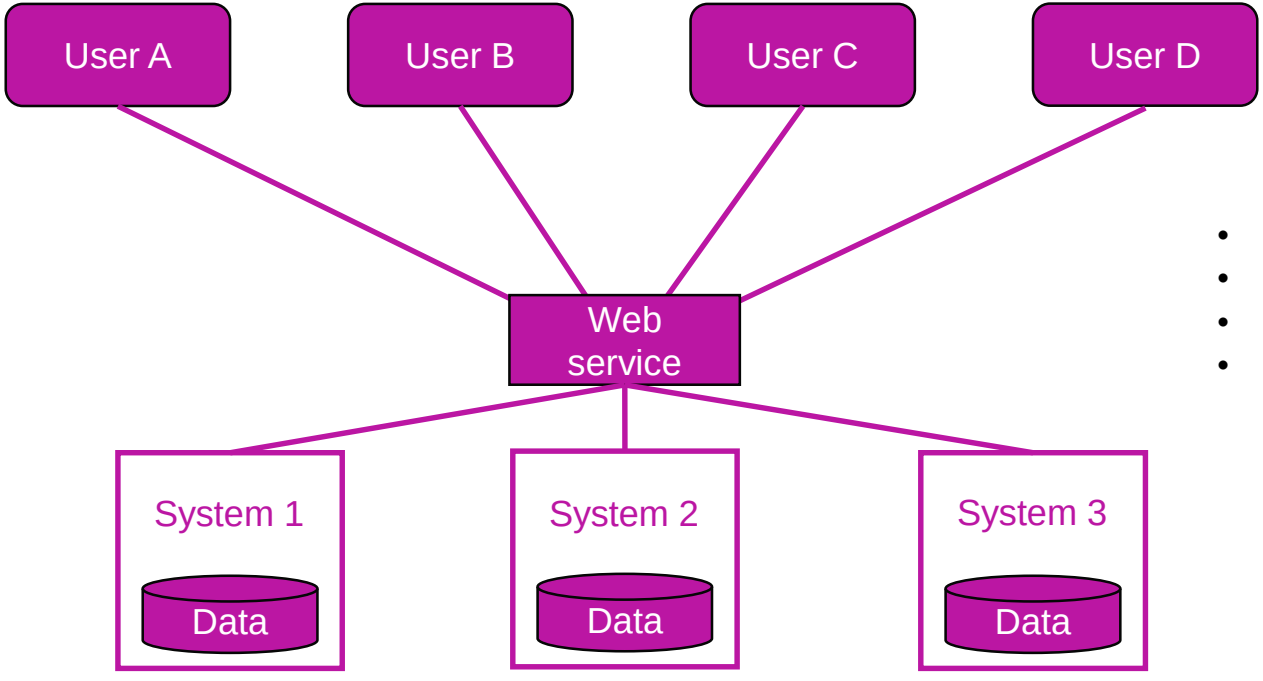
- Different Organizations
- ??? semantics
- Same syntactics
- (Foundations and Systems are not that important)

Different interoperability scenarios



- Cooperating organizations
- Same(?) semantics
- Same syntactics
- (Foundations and Systems are not that important)

Different interoperability scenarios



- Same(?) organization
- Same semantics
- Same syntactics
- (Foundations and Systems are not that important)

Different interoperability scenarios

While increased interoperability is typically beneficial, the best level of integration between different services is very much case-dependent. Sometimes increased integration can also provide a hindrance (e.g. by creating significant amount of inertia in various processes)

For the next time…

Fifth (and final!) exercise round has been published

Remember to continue writing your learning diary (also for next week's material)

Take into account the peer assessment you got