Distributed Geodatabases in Archaeological Joint Research Projects

Prof. Dr. phil. Kai-Christian Bruhn
Dipl.-Ing. Tobias Kohr
1. The projects: HiGeoMes and Eifel GIS
2. Data quality perspective
3. Re-using existing data
4. Data capture – Minimizing falsification
5. Modelling uncertainty
6. Data dissemination – Describing temporal and spatial uncertainty
7. Semantics – From description to signification
1. HiGeoMes

- Historical Geography of Upper-Mesopotamia
- Integration of textual and archaeological data
- Binational cooperation:
  - FU Berlin
  - Universität Mainz
  - FH Mainz/i3mainz
  - Collège de France Paris
  - Université de Bourgogne Dijon
- Funded by: DFG, ANR
1. Eifel GIS

- Part of GeoCycles Research Centre
- Integration of geographical and temporal information in archaeological data
- Cooperation of:
  - Universität Mainz
  - FH Mainz/i3mainz
  - Max Planck Institut für Chemie
  - Römisch-Germanisches Zentralmuseum (RGZM)
- Funded by: Rheinland-Pfalz
1. HiGeoMes and Eifel GIS

- Webportal: Tool to research and visualize archaeological sites
- Webservices: Dissemination of archaeological geodata
  - Technically standardized
  - Metadata
  - Access Control / License terms
  - Integration
- Contribution to architecture of research data infrastructures in archaeology
  - Embedding existing data
    - Maps
    - Satellite images
    - Administrative geodata
  - Open to further data sources
1. HiGeoMes and Eifel GIS

Aspects of geospatial data quality [D. Cornford, M. Williams]

- **accuracy (uncertainty):** value correctly represents the real world
- **completeness:** degree of data coverage for a given region and time
- **consistency:** are rules to which the data should conform with
- **usability:** how easy is it to access and use the data
- **traceability:** can one see how the results have arisen
- **utility:** what is the user view of the data value to their use-case

Valid and applicable for sensor data
Suitable as a starting point for describing archaeological data quality
3. Re-use of existing data

Eifel GIS

+ Ad-hoc integration of administrative geodata
3. Re-use of existing data

HiGeoMes

+ Rich digital data sources available for research area
+ Published as supplementary information and web portals

Encountered lack of:
- Usability
- Accuracy
- Consistency
3. Re-use of existing data

Usability

Bjoern Menze; Jason Ur, 2011, "Patterns of long-term settlement in the Upper Khabur Basin, northeastern Syria.“ hdl:1902.1/17731

- Data published as supplementary information
- Almost full record available (CSV, GeoTIFF)
- No interfaces / Webservices to access the data
- Proprietary CRS
  - Extensive processing
  - Dependency on specific software
3. Re-use of existing data
Usability

Corona satellite imagery, http://corona.cast.uark.edu/index.html

- High resolution imagery (6ft ground resolution)
- Documenting landscapes lost today (dam projects)
- Open license
- Proprietary (ESRI) Webservices available
3. Re-use of existing data

- **Usability**
  - **Corona** satellite imagery, [http://corona.cast.uark.edu/index.html](http://corona.cast.uark.edu/index.html)
  - High resolution imagery (6ft ground resolution)
  - Documenting landscapes lost today (e.g., dam projects)
  - Open license
  - Proprietary (ESRI) Webservices available

---

3. Re-use of existing data

Accuracy and consistency

Archaeological sites dataset from ShareGeo, http://www.sharegeo.ac.uk/handle/10672/24

- 1700 potential sites in research area
- Inconsistent data sources
- Detailed description of data
- Low spatial accuracy
3. Re-use of existing data

Accuracy and consistency

Archaeological sites dataset from ShareGeo, http://www.sharegeo.ac.uk/handle/10672/24

- 1700 potential sites in research area
- Inconsistent data sources
- Detailed description of data
- Low spatial accuracy
Accuracy and consistence

Archaeological sites dataset from ShareGeo, http://www.sharegeo.ac.uk/handle/10672/24

- 1700 potential sites in research area
- Inconsistent data sources
- Detailed description of data
- Low spatial accuracy
3. Re-use of existing data

Accuracy and consistency

Archaeological sites dataset from ShareGeo, http://www.sharegeo.ac.uk/handle/10672/24

• 1700 potential sites in research area
• Inconsistent data sources
• Detailed description of
• Low spatial accuracy
4. Data capture

Different reliable methods for non-IT-enthusiasts

HiGeoMes
Customized desktop GIS with DB connection (via Webservice)

Eifel GIS
Web GIS with Feature Transaction Service
4. Data capture

Minimizing falsification
• Semi-automatic control mechanisms
  • forms
  • controlled vocabularies
  • refinement / plausibility checks
• Low learning curve

Distributed environment
4. Data capture

Minimizing falsification
- Semi-automatic control mechanisms
  - forms
  - controlled vocabularies
  - refinement / plausibility checks
- Low learning curve

Distributed environment
Scale of accuracy based on files of official cultural heritage administration:

- Precise – coordinates, parcel
- Approximate – description
- Locality – municipality

Intentional random falsification

- To prevent looting of archaeological sites
- Service abstraction layer has been invented for such tasks
5. Modelling spatial uncertainty – Eifel GIS

- Precise: coordinates, parcel
- Approximate: description
- Locality: municipality

Intentional random falsification

To prevent looting of archaeological sites

Service abstraction layer has been invented for such tasks
5. Modelling of temporal uncertainty – HiGeoMes

Definition of periods

- MBA (Middle Bronze Age)
  - MBA I
  - MBA II
- LBA (Late Bronze Age)
  - LBA I
  - LBA II

Encoding of (un-)certainty

- 0: unknown
- 1: unsure
- 2: probable
- 3: sure
5. Modelling of temporal uncertainty – HiGeoMes

Definition of periods
- MBA (Middle Bronze Age)
  - MBA I
  - MBA II
- LBA (Late Bronze Age)
  - LBA I
  - LBA II

Encoding of (un-)certainty
- 0: unknown
- 1: unsure
- 2: probable
- 3: sure

Outlook: More flexible in relational model
6. Data dissemination

- Based on existing standards and concepts (OGC, W3C)
- Using implemented mature technologies (PostgreSQL/PostGIS, MapServer, OpenLayers, GeoExt, MapFish)
- Most archaeological data comprise spatial information
- INSPIRE includes protected sites

⇒ Following the approach of Spatial Data Infrastructures
HiGeoMes sites are published via Web Feature Service (WFS)

- GetCapabilities: Information about service and data
- DescribeFeatureType: Description of data-structure
- GetFeature: Requesting data
6. Data dissemination

HiGeoMes sites are published via Web Feature Service (WFS)

- GetCapabilities: Information about service and data
- DescribeFeatureType: Description of data-structure
- GetFeature: Requesting data
HiGeoMes sites are published via Web Feature Service (WFS)

- **GetCapabilities**: Information about service and data
- **DescribeFeatureType**: Description of data-structure
- **GetFeature**: Requesting data

HiGeoMes sites are published via Web Feature Service (WFS)

- GetCapabilities: Information about service and data
- DescribeFeatureType: Description of data-structure
- GetFeature: Requesting data

6. Data dissemination

HiGeoMes sites are published via Web Feature Service (WFS)

- GetCapabilities: Information about service and data
- DescribeFeatureType: Description of data-structure
- GetFeature: Requesting data

→ Various options to describe temporal and spatial uncertainties
GetCapabilities

• Layer - Abstract

- <FeatureType>
  <Name>Higecmes_Sites</Name>
  <Title>Higecmes Sites</Title>
- <Abstract>
  This layer provides archaeological sites from the middle and late bronze age from Upper Mesopotamia, which were collected within the HiGeoMes project funded by DFG and ANR. Data were collected and assessed by Christoph Fink, Institut für Ägyptologie und Altertumswissenschaft, Johannes-Gutenberg-Universität Mainz from available published archaeological reports. The datasets comprise a HiGeoMes-ID as primary key. Where appropriate the ID of the ANE list (http://www.lingfil.uu.se/staff/olof_pedersen/Google_Earth/, accessed 11/2011) is referred to. The dating of sites is organized according to the following terms: MBA = Middle Bronze Age and LBA = Late Bronze Age, divided into sub-phases I and II each. These confidence of dating is encoded from 0 to 3: 0-unknown, 1-unsure, 2-probable, 3-certain
</Abstract>
<DefaultSRS>urn:ogc:def:crs:EPSG::4326</DefaultSRS>
- <OutputFormat>
  <Format>text/xml; subtype=gml/3.1.1</Format>
</OutputFormat>
- <ows:WGS84BoundingBox dimensions="2">
  <ows:LowerCorner>37.865 34.475</ows:LowerCorner>
  <ows:UpperCorner>43.979 38.219</ows:UpperCorner>
</ows:WGS84BoundingBox>
<MetadataURL format="text/html" type="TC211">http://metadata.higecmes_sites.html</MetadataURL>
</FeatureType>
6. Data dissemination

GetCapabilities

• Layer - Abstract

    <FeatureType>
    <Name>Higeomes_Sites</Name>
    <Title>Higeomes Sites</Title>
    <Abstract>
    This layer provides archaeological sites from the middle and late bronze age from Upper Mesopotamia, which were collected within the HiGeoMes project funded by DFG and ANR. Data were collected and assessed by Christoph Fink, Institut für Ägyptologie und Altertumswissenschaft, Johannes-Gutenberg-Universität Mainz from available published archaeological reports. The datasets comprise a HiGeoMes-ID as primary key. Where appropriate the ID of the ANE list (http://www.lingfil.uu.se/staff/olof_pedersen/Google_Earth/, accessed 11/2011) is referred to. The dating of sites is organized according to the following terms: MBA = Middle Bronze Age and LBA = Late Bronze Age, divided into sub-phases I and II each. These confidence of dating is encoded from 0 to 3: 0-unknown; 1-unsure; 2-probable; 3-certain
    </Abstract>
    <OutputFormat>
    <Format>text/xml; subtype=gml/3.1.1</Format>
    </OutputFormat>
    <ows:WGS84BoundingBox dimensions="2">
    <ows:LowerCorner>37.885 34.475</ows:LowerCorner>
    <ows:UpperCorner>43.979 38.219</ows:UpperCorner>
    </ows:WGS84BoundingBox>
    <MetadataURL format="text/html" type="TC211">http://metadata/higeomes_sites.html</MetadataURL>
    </FeatureType>
6. Data dissemination

GetCapabilities

- Layer - Abstract

- <FeatureType>
  - <Name>Higeomes_Sites</Name>
  - <Title>Higeomes Sites</Title>
  - <Abstract>
  This layer provides archaeological sites from within the HiGeoMes project funded by DFG; Archäologie und Alterorientistik, Johannes-C.
  datasets comprise a HiGeoMes-ID as primary, encoded from 0 to 3: 0-unknown, 1-unsure, 2-MBA = Middle Bronze Age and LBA = Late Bronze Age.
  - <DefaultSRS>urn:ogc:def:crs:EPSG::4326</DefaultSRS>
  - <OutputFormat>
  - <Format>text/xml; subtype=gml/3.1.1</Format>
  - <OutputFormat>
  - <ows:WGS84BoundingBox dimensions="2">
  - <ows:LowerCorner>37.863 34.475</ows:LowerCorner>
  - <ows:UpperCorner>43.979 38.219</ows:UpperCorner>
  - <ows:WGS84BoundingBox></ows:WGS84BoundingBox>
  - <MetadataURL format="text/html" type="TC211">http://</MetadataURL>
  - <FeatureType>

- Layer - MetadataURL
DescribeFeatureType

- Definition of custom data type and provision of its schema

```xml
<extension base="gml:AbstractFeatureType">
  <!-- sequence -->
  <element name="imqGeometry" type="gml:GeometryPropertyType" minOccurs="0" maxOccurs="1">
    <element name="jklf" type="string"/>
  </element>
  <element name="ANE-ID" type="integer"/>
  <element name="Modern_Name" type="string"/>
  <element name="Name_MBA" type="string"/>
  <element name="Name_LBA" type="string"/>
  <element name="Description" type="string"/>
  <element name="Dating_MBA_I" type="integer"/>
  <element name="Dating_MBA_II" type="integer"/>
  <element name="Dating_LBA_I" type="integer"/>
  <element name="Dating_LBA_II" type="integer"/>
  <element name="Dating_LBA_III" type="integer"/>
</extension>
</complexContent>
</complexType>
</schema>
```
DescribeFeatureType

- Definition of custom data type and provision of its schema

Possibility to use existing data types
6. Data dissemination

DescribeFeatureType
- Definition of custom data type and provision of its schema

Possibility to use existing data types or extend complex data type by a custom definition
6. Data dissemination

DescribeFeatureType

- Definition of custom data type and provision of its schema

New application schemas contribute to standardization process

or extend complex data type by a custom definition

Source: Fernández Freire et al.
6. Data dissemination

Re-use of published data
- Archaeological sites
- Topographical maps
- Satellite imagery
- Terrain model
Conclusion:
• OGC services are suitable to describe data on a syntactical level

But: How can the data‘s signification be disseminated?
⇒ Integration with other technologies

Current research topics:
• Semantic modelling (ontologies)
• Linked data
7. Semantics – From description to signification
7. Semantics – From description to signification
References

Dan Cornford, Matthew Williams, “Uncertainty and Quality UncertML, UncertWeb and Geoviqua”, OGC meeting, Bonn, 03/2011


Bjoern Menze; Jason Ur, 2011, "Patterns of long-term settlement in the Upper Khabur Basin, northeastern Syria." hdl:1902.1/17731


Antone J. Mathys, “A GIS comparative analysis of bronze age settlement patterns and the contemporary physical landscape in the Jazira region of Syria”
Questions?

Prof. Dr. phil. Kai-Christian Bruhn
bruhn@geoinform.fh-mainz.de

Dipl.-Ing. Tobias Kohr
tobias.kohr@geoinform.fh-mainz.de