**Summary of basics**

Project aim for this work package:

* Creation of open online database hosting Roman road data.
* Data can be explored, queried, downloaded, added and edited by registered users.

Our short-term urgent basic requirements for autumn 2021:

* Online database with road segments as core entities
* Hosted on secure backed up space accessible by project collaborators only
* Ability to upload shapefiles in which rows represent road segments
* Automatic mapping of attributes of road segments to relevant fields in database
* Visualisation of road segments in database on webmap
* Map-based query of data to select, visualize, export subsets
* Avoid need for additional steps when uploading data collected offline to online data management platform

Our ideal longer-term desirables are different, see c-d.

**Supplementary information:**

1. Current itiner-e database <http://itiner-e.org/index.php/crud/routes>
2. ER diagram of current database (not representative of current minimum short-term desirables): ‘ER.png’
3. Sample shapefile of road segments from Gallia (France): ‘Gallia3.zip’

**Relevant other projects:**

* Pleiades: this is a gazetteer with placenames with stable URI’s per placename. Our project’s minimum final requirement is comparable: an open online database with stable URI’s for its entities <https://pleiades.stoa.org>
* Viabundus: Pre-modern route planner. Ideally at a very late stage (2023-24?) we would have a graph representation and path-based query functionality like this <https://www.landesgeschichte.uni-goettingen.de/handelsstrassen/index.php>
* Orbis: Roman route planner <https://orbis.stanford.edu>
* Recogito a way to relate roads, stops (mansiones) and Pleiades URIs <https://recogito.pelagios.org/>

1. **specifications**

*break down your data: what is the smallest unit(-s) for points and distances, e.g., geom distance between two places, potsherd*

* Basics:
  + The core entity in our current urgent phase of data collection is a route segment (represented as a row in the .shp file).
  + Spatially the smallest points represent a centre point of a settlement/village/villa/fort, a crossroads, a bridge, a turn in a road (see polylines in .shp). These are considered features of the geometry and not the key entities.
  + Spatial focus of data collection is winding roads between settlements, not roads within settlements. The spatial scale is on the level of landscapes and not on the level of individual settlements.
  + Spatial precision is roughly 500m
* Later development:
  + The spatial data collected in the current urgent phase (represented by .shp) should be querrieable on the level of units such as any attribute values (e.g. created by, date, bibliography, rows), or any spatial geometry (places with or without Pleiades URI’s, road segments).
  + Spatial data should be transformable into graph data. Nodes represent places (settlements, crossroads, bridges, villas, forts), edges represent the ability to move from one node to another node over a road documented in the database without passing a third node.

*your database's relationship to pleiades.*

* Basics: none (i.e. the current urgent phase of data collection does not include sufficient time to enter Pleiades ID’s to the data).
* Later development:
  + each route segment goes from an ancient town to another ancient town (see ‘name’ column in .shp attribute table). Fields need to be added to the route segment table that include Pleiades\_from\_uri and Pleiades\_to\_uri (1 to 1).
  + Each route segment goes through a set of ancient towns. A set of Pleiades URI’s should be linkable to each route segment (1 to many).

*relationship to extracted data*

(I am not sure what you mean by extracted data)

End-users must be able to download our data to reuse it in their own projects. Specially the geographical paths of the roads and at least some basic information: name, date, author, bibliography,…

1. **usecases**

*what will your researchers be doing (****in the fall****)*

* Digitization of road segments:
  + Happens offline, not directly in the online database
  + Data managed during collection offline in .shp file
  + Each row represents a road segment from an ancient place to another ancient place
  + Each row includes attribute information about the road segment
  + Creation of road segments motivated by segments that do not include differences in attribute information (e.g. the attribute information for row section is representative of the data collected for the entire spatial geometry of that route segment: same type, date, bibliography).
* Uploading .shp files in batches to online database:
  + To safely store finished subsets of the data collection
  + To enable visual overviews of own data collection
  + To allow all project members to see progress made by others
  + To allow all project members to error-check spatial and tabular data representations
* Downloading the data to keep backups of all the roads uploaded in a single file
* Performed by Pau de Soto (west) and Adam Pažout (east)

*how will the end-users be using database+website*

Later development:

* An open online community resource launched to the public in 2023-2024:
  + Everyone who visits the website has view, query and download rights to the data.
  + A map-based interface to display the total spatial extent of the Roman Empire and spatial dataset (e.g. see Orbis).
  + Map-based selection/query functionality to explore subsets of data as spatial or tabular representations that can be exported as .csv for tabular data or spatial data formats for geometry (.kml, .geojson).
  + Text-based selection/query functionality to search for subsets of data related to specific placename (e.g. Rome) or named route (e.g. Via Augusta). Export functionality of results.
  + Field-based query of attributes: e.g. within date range X-Y, in Roman province X, from bibliography X. Export functionality of results.
  + Users can optionally choose to register, are vetted by project, and granted login access to give them upload, edit, archive rights (maybe not delete but place in an archive to enable restoring). Edits made by these users with these rights should be identifiable, such that errors can be restored.
  + Graph data representation of entire dataset curated by project members (referred to as canonical network). Available for download.
    - This requires project members to specify for the case of the canonical network which places Pleiades URI’s are connected to which route segments and in what order.
  + Optional because dependent on time constraints (priority is given to spatial data collection):
    - Graph-based query of canonical network: fastest path over network from place A to place B. Distance, financial cost, effort in calories, modes of transport, travel time, will all be pre-calculated by project team for the case of the canonical network only.
    - Make custom graph data representation of subset of spatial dataset by end-user representing the end-user’s assumptions for the conditions to connect a pair of nodes. E.g. “give me the network of those roads in the database that I find particularly reliable”. Basically a transformation into graph data from a spatial data format of a set of route segments and a set of places (latitude longitude sets with attribute information describing ancient name, Pleiades URI, site type etc.), following a set of assumptions for network data creation (e.g. places within 500m of a road segment are considered to be on the road, isolated places are connected to the nearest crossroads). This could take the form of a piece of software that can be downloaded and performed offline on data exported from the database.

*how will you be using the database*

In addition to answers in (b) above:

* Short-term:
  + exporting of subsets for creation of research-context specific spatial and network data used to test hypotheses in GIS and simulation models
* Long-term:
  + management of resource to ensure URI’s remain stable, data remains accessible.
  + correction and refinement of existing data.
  + addition of new data to database.

1. **prioritization**

*formulate need-to-haves and nice-to-haves*

Need-to-haves:

* Use of itiner-e.org domain name
* Long-term stability of resource
* Stable URI’s for entities
* Places linkable to Pleiades URI’s
* Online database with road segments as core entities
* Hosted on secure backed up space accessible by project collaborators only
* Read/write access by project members
* Ability to upload shapefiles in which rows represent road segments
* Automatic mapping of attributes of road segments to relevant fields in database
* Visualisation of road segments in database on webmap
* Map-based query of data to select, visualize, export subsets
* Avoid need for additional steps when uploading data collected offline to online data management platform
* Map-based, text-based and field-based query of data.
* Export of subsets in spatial and tabular formats.

Nice-to-haves:

* Canonical network that can be querried (i.e. route planner)
* Transformation of subsets into graph data
* Exciting interactive GUI

1. **outcome and ultimate goal**

*are we solving multiple small tasks (easier upload, graph database for spatial analytics...)*

*OR one generic solution*

Currently it feels more like multiple small tasks: support data collection, project’s online database management system, gazetteer for end-users, querry functionality, route planner on canonical network.

*final website, how does it work, what does it looks like*

* Intuitive welcome page that is visually appealing (I will use other budgets for that) and encourages exploration
* Looks and works like a combinations Pleiades and Orbis with a focus on the former: This is an authoritative well-documented gazetteer that can be searched and cited; which offers the further option of route-finding in a canonical network.