

# ATHENA

## Access Tool for data on Historical Ecology and ENvironmental Archeology

Jan Luiten van Zanden (UU)

Rob Lenders (RU)

Joop Schaminée (WUR)

Stephan Hennekens (Alterra)

Theo Spek (RUG)

Thomas van Goethem (UU/RU)

### Summary for general public

#### **Mens-natuur relaties in historisch perspectief**

In de ATHENA-data portal wordt de historische mens-dier relatie vastgelegd. Daarbij wordt een breed scala aan bronnen gebruikt: van archeologische data tot schilderijen. De data portal vormt een uniek uitgangspunt voor historici, archeologen en biologen om de historische relatie mens-natuur te onderzoeken.

**Keywords:** Human – nature relations; Historical biodiversity; Interdisciplinary research; Research infrastructure; Database innovations

## **Contents:**

### **1. Executive summary**

### **2. Research programme and research field(s)**

- **Human-nature relationships**
- **Originality and innovative nature of the research program**

### **3. Description of the data model**

- **ATHENA: the strategy**
- **Design of the central data portal**
- **Development and technical feasibility**

### **4. Types of datasets and their feasibility**

- **Data types and contexts**
- **Historical data**
- **Archeological data**
- **Biological/observational data**

### **5. Partners and consortium**

### **6. Planning and work packages**

- **Work packages**
- **Planning**

## 1. Executive summary

ATHENA envisions the creation of a data portal that will hold information on historical contexts of human – nature relationships for a broad variety of plant and animal species and the landscapes and ecosystems they live(d) in. Aim of the ATHENA project is to develop an internationally unique data portal allowing researchers from multiple scientific disciplines to study human-nature relationships in an interdisciplinary environment.

Studying the processes in which people transformed their surroundings and the ways in which the natural world has influenced the course of human history is important to understand the current state of human-nature relationships. Such relationships can best be studied on the basis of comprehensive temporal (and geographical) explicit datasets. Currently, however, data and datasets are scattered or even lacking. Moreover, the data that is available is generally missing historical, social, cultural and ecological context. To address this situation, historians, archaeologists and ecologists must come together to share and combine data. The goal of the ATHENA project is to produce and maintain a data portal that will provide a platform for large-scale, comparative (both in space and time) and multilevel studies of human-nature relationships.

In ATHENA a wide range of data sources from the fields of history, archaeology and ecology will be brought together in an access portal. The portal will be developed by means of which mutually linked subdatabases can be approached and information can be combined and integrated. Databases on flora and fauna containing geographically explicit observational and archaeological information, and historical databases containing, or referring to historical written, pictorial and oral sources will be linked in the confines of ATHENA. The existing databases will also be enriched with relevant information which has not yet been digitized before. Therefore, an entirely new database will be developed for socio-economic and cultural data related to human-nature interactions, as well as a thesaurus for a set of selected animal and plant species. By applying state of the art tools these large data files can be integrated and queried resulting in meaningful content. Furthermore, to facilitate inclusion of historical data not yet made accessible in any database, external online input portals are to be developed or adapted. These input portals will enable scientists and other professionals, but also amateur historians and ecologists, to deliver new data to the system. The ATHENA data portal is a unique starting point for historians, archaeologists and biologists to investigate major questions in the field of human-nature relationships in an interdisciplinary environment.

To accomplish this project, a broad consortium has been brought together. ATHENA is a joint project of a core team of scientists, supported by a consortium of knowledge and heritage institutions and public organizations. University partners include Radboud University, University Utrecht, VU University Amsterdam, Wageningen University and the University of Groningen. Coalition partners from the humanities include DANS (data archiving), the Meertens Institute (texts), RKD (iconographic material) and RCE (plant and animal residues). Partners from the natural sciences include SOVON (bird observations), FLORON (plants), RAVON (reptilians, amphibians and fish), Mammal Society (mammals) and Alterra (ecology / vegetation).

## **2. Research programme and research field(s)**

### **Human-nature relationships**

The various dialogues over time between people and the rest of nature have led to mutual impacts and can be characterized by complex interactions. Socio-cultural activities and processes are not only dependent on the ecological system, they have also far reaching impact on the ecological system. In the course of time human impact on the natural world has substantially increased. To understand these impacts we must try to unite the worlds of natural sciences and humanities. An interdisciplinary view on the historical relationship of man and nature enables scientists to study long-term interrelationships between socio-cultural and ecological features. Aim of this project is to facilitate the exploration of the ways in which the natural world has influenced the course of human history and the ways in which people have thought about and transformed their surroundings. To that end, collecting, combining and integrating information from different historical sources is needed. The resulting information pool, or data portal, can facilitate a broad spectrum of research.

Human – nature interactions can be addressed in several lines of research. The first deals with the socio-economic aspects. What is the influence and impact of nature – the biophysical world – on politics, policies and the economy? Studies of pre-modern economy, politics, culture, food preparation and consumption in relation to nature, show that changing environmental conditions have had a profound effect on human history. The impacts of resource scarcity on politics, economy and society, for instance, should be studied in long term and multilevel analyses. The second line of research revolves around the natural realm. The natural world is not static, it reacts on our actions. Human impact on nature has a wide spectrum of consequences for ecological systems. By reconstructing historical ecosystems the effects of current and historical levels of human impact can be assessed. Biological baselines, corresponding to ‘pristine’ or ‘minimally disturbed’, aid nature conservationists to develop holistic, environmentally and historically valid plans of action. The historical perspective also permits biologists to determine the extent to which ecosystems can deal with environmental pressure without crossing biodiversity thresholds. The third line of research deals with the human intellectual realm consisting of perceptions, ethics, myths and other mental constructs related to the natural world. To understand what people think about nature and how they have expressed those ideas in folk religions, popular culture, literature and art. Either directly or indirectly, cultural phenomena have influenced nature. Ideas from the human intellectual realm about the biophysical world influence the way we deal with the natural environment. This involves, amongst others, research on the direct influence of changes in consumption and production patterns, food collection and agricultural practices but also of the impact of ideological, religious and philosophical viewpoints on biodiversity development.

The human – nature relationship consists of bilateral interactions, so the lines of research are not to be considered as separate fields, but should be interlinked. The combination of ecological, historical and archaeological research will lead to a more complete comprehension of the historical relationship of man and nature. The ATHENA project facilitates advancements in these broad fields of research and aims to integrate the individual sciences involved. This approach makes history more inclusive in its narratives by examining the

historical role and place of nature in human life and improves conceptual and mathematical models of human impacts on nature.

### **Originality and innovative nature of the research program**

An interdisciplinary approach is necessary to fully understand the impacts and consequences of human-nature dialogues. By collecting, combining and integrating information from different historical sources, scientists, policy makers and the broader public can more easily study and articulate changes in mutual relationships between socio-cultural and ecological systems. Rather than concentrating on one specific event in time and space, ATHENA allows to study and understand human-nature interactions across temporal and spatial scales in order to gain a full understanding of its cumulative effects. By working with data from different disciplines information quality can be improved significantly, as various datasets can be calibrated where there is a temporal overlap in archaeological, historical and observational data. In addition, information on historical, social, cultural and ecological phenomena provides much needed contexts for human-nature relationships research. Currently, however, the available information on historical presence of species, on species assemblages and on human – nature relationships from the majority of sources is scattered or poorly opened up. Moreover, the data that is available is generally missing historical, social, cultural and ecological context. The ability to create such comprehensive temporal and geographical (explicit) datasets, as proposed in the ATHENA project, will therefore greatly benefit the disciplines involved.

Combining information from different historical sources can enable scientists to study human – nature relationships in a way that would be far more difficult or even impossible by the use of information from only one type of source. For archaeologists, additional historical information from written, oral and pictorial sources can help in interpreting archaeological findings. Such additional information can therefore help in understanding how human society did cope with then-prevailing environmental and cultural circumstances. For historians, evidence on the historical presence of species inferred from natural-historical sources and archaeological data can contribute to describing and understanding historically developments in (socio-) economics and land-use, but also in jurisdiction, popular tradition, medicine, art and religion. For ecologists and environmental scientists, both archaeological and historical information greatly enhances the understanding of developments in biodiversity, long-term ecological and biogeographical processes and human influences thereon, and reconstructing historical references for ecological rehabilitation.

Summarizing, the ATHENA project will develop a data portal in which archaeological information, observational data and historical sources on flora and fauna are brought together and combined. The data portal provides much needed additional and contextual information to the individual sciences involved. By doing so the project alerts disciplinary scientist to novel data sources that they are unfamiliar with or not even aware of. This is the true innovative feature of the ATHENA project, combining and ultimately integrating scientific disciplines through data. The interdisciplinary nature of the project will greatly enhance the understanding of human-nature relationships, an endeavor not likely to be achieved by a single discipline alone. The internationally unique database will offer the scientific community as well as professionals from governmental and (environmental) management agencies a unique tool for analyzing historical human-nature relationships.

### **3. Description of the data model**

#### **ATHENA: the strategy**

ATHENA builds on the model of data hubs. Data hubs offer data collections on specific issues. The project aims to create a network of databases that provide state-of-the-art data for their own fields of research, and that are linked in the joint effort to create a body of knowledge on historical human-nature relationships. The datasets presented in ATHENA are collected and selected by teams of researchers supported by technology. The proposed online data portal gives either direct historical information on species and/or references to relevant sources. ATHENA, however, is much more than just a portal to data collections. It aims to combine data, harmonize metadata, develop standards and ontologies, and provide state-of-the-art visualizations. All these different functions of an innovative research infrastructure will be developed within a community of researchers. In this community the researchers involved bring the know-how needed to develop and maintain data collection for their respective scientific fields, whilst also acquiring access to novel data sources with which they can enrich their research. Through collaborations standard formats and a common 'research language' can be developed needed for true interdisciplinary research on human-nature relationships.

Developments in information sciences have enhanced the possibilities in data collection through digitization and database development. New methods of data-sharing and integration are emerging and need to be exploited. Through these new technologies ATHENA enables existing archaeological, ecological and historical databases to be brought together. ATHENA will also make many valuable historical data on human-nature relations that are not yet included in existing databases available. Cross-domain interoperability and comparability of large datasets is one of the main objectives of the ATHENA project. Additionally, advancements in data mining techniques will provide a tool to swiftly identify and select relevant data from large collections of unstructured or semi-structured historical sources. Another important technical innovation is the development of infrastructure for citizen science as it is a way to greatly expand the existing databases. All these different functions of an innovative research infrastructure will be developed not for but within a community of researchers. The partners and institutions in the ATHENA consortium have ample experience with database management, data mining and citizen science.

#### **Design of the central data portal**

Central in ATHENA are 'species', meaning not only species in a strict biological sense (indicated by a scientific genus and species name) but also units of higher and lower taxonomic order (Class – Order –Family vs. subspecies and varieties), and even mythical species like basilisks, werewolves and unicorns. Species thus form the main entrance of ATHENA. From this entrance subdatabases are queried, yielding information from historical, archaeological and ecological sources. The portal can be consulted on the following (preliminary) thematic fields; nomenclature, non-fiction literature, art and culture, archaeology, natural history, ecology, socio-economics (Figure 1). Mutual data streams are foreseen between the respective fields. For instance, in each thematic field geographically explicit information and information of a more

generic nature can be found. Relevant geographical information on changes in distribution of species may be inferred from each thematic field, and combined.

The end goal is to make a data portal that:

- Harvests data from existing databases.
- Harmonizes the data in such a way that they can be presented with a uniform structure and format.
- Enriches the data with appropriate geographical and temporal contexts.
- Provides a tool for selection, combination and aggregation of data according to taxonomy, geographical range and time lines. Through this tool users are able to construct and download datasets.
- Stores retrieved data and makes them available for further processing and research.
- Provides a shell for presentation and visualization.

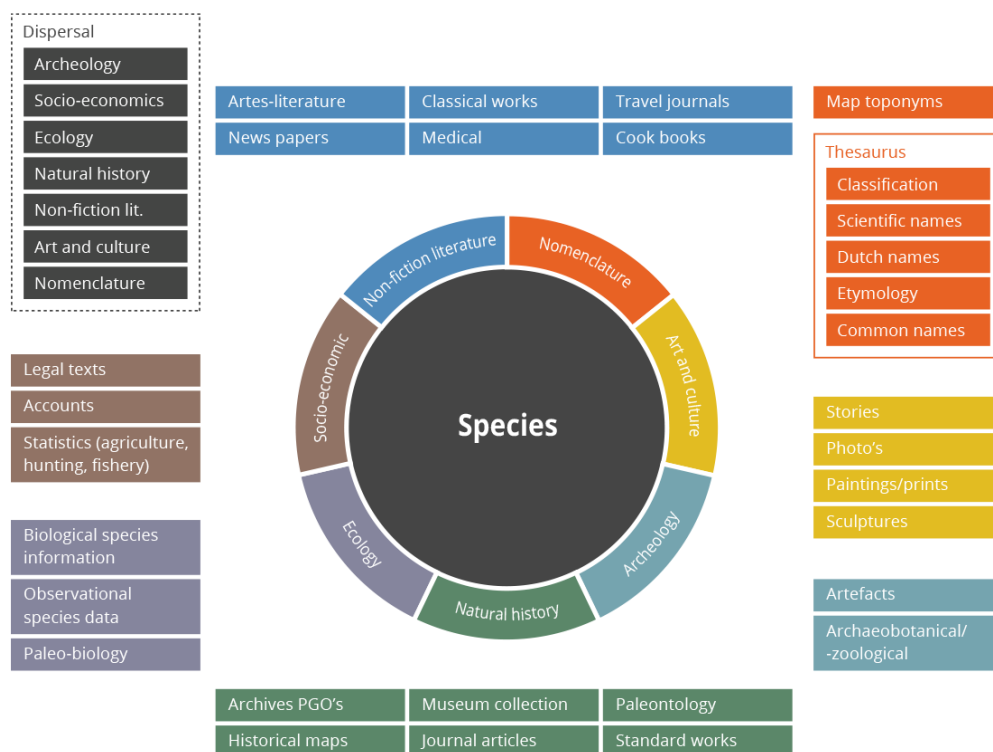


Figure 1. Data sources divided in thematic fields

### Development and technical feasibility

The ATHENA data portal is an ambitious project as different data types are brought together, including textual, pictorial, (semi-) structured and unstructured data from three fields of research, i.e. historical, archaeological and ecological. The heterogeneous nature of the different data hubs/datasets pose several challenges concerning properly linking and querying the data. First, the respective datasets will be linked in the ATHENA data portal through APIs provided by the respective databases. Simply linking the respective databases, however, is not sufficient. The data portal has a user-interface for querying different data sources (databases) and constructing and downloading composite datasets. A general concern is the fact that the

information is heterogeneous in the way it is related to the species. For instance, the archaeological databases contain geographically and temporal explicit information, while the historical database may have datasets concerning food collection in general. The data available thus differs between species and data types. The various data types have different characteristics and are therefore not published in an uniform manner, with some data sets even lacking meta-data or semantic annotations. This makes it difficult to design an uniform format for metadata and develop an domain specific ontology that is applicable to all species and data types. Computational tools will be used to properly link the different datasets. The data will be published as Linked Data through the use of semantic web technology such as the RDF format, with automatic RDF metadata generation as an option. Furthermore, provenance information, such as information on data transformation, origins of the data and which datasets were used to compute the query, will be made available with the query results. This is important as the end user of the ATHENA data portal needs to have information on/and access to the 'raw' data.

ATHENA will be realized through several incremental steps. First, data models of existing databases, developed and managed by successful communities of practice in the fields of historical, archaeological and ecological research, will be identified and subsequently integrated and mapped on the ATHENA data model. The heterogeneous data sources are then harmonized to be able to appropriately search and combine the different databases within the confines of ATHENA. Next, data integration is taken one step further through the use of new technologies. The aim is to create a central data portal with an interface that presents the dynamic historical data in a clear and attractive way. Semantic search algorithms will be developed to study human-nature relationships through representative formats. Furthermore, novel visualization tools will allow time and geography scoped views on the presented data.

The following (detailed) development steps can be distinguished:

#### 1. Data model development

- Determine model requirements: a case study with several key species
- Identify existing data models and extend these where necessary
- Describe data models of provided data sets and map these models to our ontology

#### 2. Data entry and development

- Populate data model with provided data sets through model driven data integration of (semi) structured data
- Evaluate data extraction techniques such as semantic annotation recognizing and natural language processing tools for automatic extraction of data from (online) text data
- Development of online crowd sourcing modules

#### 3. Query builder logic

- Allow users to build sophisticated queries that select particular time slices as well as allow keyword search
- Develop scalable semantic search algorithms that select contextual information based on relations between source and the user provided topic

#### 4. Visualization dashboard

- Develop spatiotemporal visualization tools for query results



In the context of the ATHENA project a large, complex data hub, a data entry system and an on-line search facility will be developed. The partners of the consortium have the competences and experience needed to efficiently build the proposed infrastructure, i.e. handling large datasets (RU, UU, PGO's), data archiving (DANS, RCE, PGO's), data validation (PGO's) and data collection (VU, RCE). Because of the unique situation as far as existing data sources are concerned, the project team will be able to relatively swiftly build the proposed data set. The development of a system for crowd sourcing – the design of an input portal and a validation system – will allow for an extension of the dataset by amateur and professional ecologists and historians.

#### **4. Types of datasets and their feasibility**

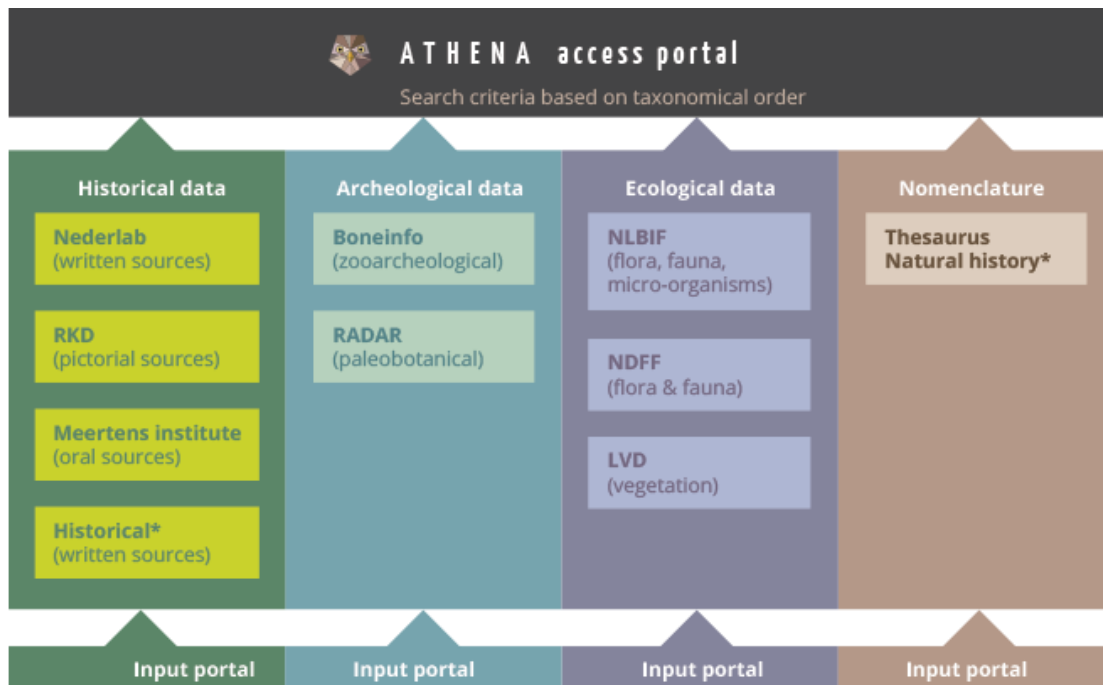
##### **Data types and contexts**

ATHENA will start to link already existing databases from three field of research, i.e. historical, archaeological and ecological (observational) sources (Figure 2.). The data from these different fields of research can be classified in different datatypes, i.e. (semi-) structured data, textual sources or pictorial (media) information.

The archaeological (BoneInfo and RADAR) and ecological (PGO's, LVD, NLBIF) databases contain structured geographically species observations. These observational records are often enriched with information on the environmental and cultural contexts. The archaeological and ecological databases are explicitly confined to the Netherlands.

The historical databases (Nederlab, RKD, Meertens institute) include newspaper archives, oral reports and pictorial sources. Some of these data sets might contain explicit geographical and temporal information on species. Most data sources, however, reveal more generic information on changes in consumption and production patterns, food collection and agricultural practices. These data sources may cover a broader geographical range since the applicability of such information in understanding the historical significance of human-nature relationships may exceed historical and present-day political borders. The preliminary geographical borders for such information are therefore set to north-western Europe.

The existing databases will also be enriched with relevant information which has not yet been digitized before. Therefore, an entirely new database will be developed for socio-economic and cultural data related to human-nature interactions. Data sources that will be included in the historical database are legal records, economic accounts, agricultural statistics, herbals, bestiaria, educational sources, travel journals and historical maps. Furthermore, a thesaurus for selected animal and plant species will be developed to successfully query historical texts.



**Figure 2. Existing and to be developed databases**

### Historical data

'Historical' data refers to information obtained from written, oral and pictorial sources pre-dating the 'present and is used as a generic term. Historical data are available in existing databases. Pictorial and oral data are available at the Netherlands Institute for Art History (RKD) and the Meertens Institute. Existing databases containing textual sources include digitized newspaper archives, book collections and journals (available at the Royal Library, DBNL, Huygens ING, INL, Meertens Institute). Through close collaboration with Nederlab these text can be searched on meaningful content. Nederlab aims to bring together all digitized texts relevant to the Dutch national heritage, the history of Dutch language and culture (c. 800 - present) in one user-friendly and tool-enriched open access web interface, allowing scholars to simultaneously search and analyse data from texts spanning the full recorded history of the Netherlands, its language and culture.

Although (written) historical data is available in digitized form and accessible for online linking and querying, many sources on historical human-nature relationships are missing. Important sources of information are not yet digitized and may contain data that can be geo-referenced, resulting in (semi-) structured data, or, provide important contextual information. These written sources include 'natural-history' works such as bestiaria and – from modern times onwards – more scientific and comprehensive books and articles, but also modern natural history magazines, e.g. De Levende Natuur, Natura, Natuurhistorisch Maandblad. Also, non-scientific historical treatises such as popular works on nature, hunting and fishery placards, market statistics, toll records, city accounts and cookery books are not yet included in any database. Such treatises do not only yield information on species in their ecological context but also concerning their role in cultural processes and activities such as economy, religion, jurisdiction and popular tradition. To make the information in these sources available an comprehensive exploratory study is needed. This inventory holds all necessary information

about the available sources and the whereabouts. The next step is to link the (digitized) sources to the relevant existing databases, or, in the case of the (semi-) structured data, include the information in the to be developed historical database.

### **Archaeological data**

The archaeological records are managed by the 'Rijksdienst voor het Cultureel Erfgoed' (RCE) and divided in two databases, for archaeobotanical and –zoological records. The archaeozoological information is available and relatively easily accessible by means of the on-line database BoneInfo. BoneInfo contains meta-information on geographical position of the excavated site, dating, type of complex, method of collection and identified species within the site. Furthermore, references are given of publications in which the findings are described and elucidated in more detail. Up to present, BoneInfo contains 1642 descriptions of sites with their respective fauna assemblages. The majority of the data concern vertebrates, but especially molluscs and to a lesser degree arthropods are also included. The archaeobotanical data is available in RADAR (Relational Archaeobotanical Database for Advanced Research). RADAR contains geographical data and research data about botanical macro remains collected during archaeological excavations. Combined, this data provides insight on the landscape, vegetation and eating habits of people in the past. RADAR contains over 175.000 sample records, collected at 1050 different sites.

### **Biological/observational data**

Historical observations of flora and fauna species are gathered in three existing databases, i.e. the National Database on Flora and Fauna species (NDFF), the Dutch National Vegetation Database (LVD) and the Dutch Biodiversity Information Facility (NLBIF). The NDFF is set up to provide information on the 'present' distribution of flora and fauna species in the Netherlands and contains in total over 40 million data. The number of historical observations, i.e. before 1950, can be found in over 8000 records. The NDFF is constantly extended by professional and amateur biologists via the online data import portals [telmee.nl](http://telmee.nl) and [waarneming.nl](http://waarneming.nl). The LVD contains vegetation descriptions of the Netherlands. In this archive 600.000 recent and historical vegetation descriptions are gathered, including the cultural and (semi-) natural landscapes. NLBIF publishes standardized online biodiversity data through the GBIF infrastructure. In GBIF 13 million records on flora, fauna and microorganisms from the Netherlands and the territorial waters are freely accessible. Important datasets include the (water)ecosystem inventarisatie of STOWA and fish data of 'Sportvisserij Nederland'. The regional museums provide valuable datasets such as regional herbaria. In total, 29 countries provide data on the Netherlands.

## **5. Partners and consortium**

The primary goal of ATHENA is to develop infrastructure that allows researchers from the humanities and natural sciences to do research. Hence, the institutions where these researcher are working are 'natural' partners in the ATHENA project. The project, however, aims to not only tend to the needs of these traditional research institutions, but also serve the 'users' of human-nature research: heritage institutions and public institutions such as nature conservation organizations, university libraries and environmental consultancy.

A strong interdisciplinary team of leading scholars and scientific institutions in the field of human-nature research has already committed to the ATHENA project. The development of the ATHENA database fits the research strategy and competences of each partner and institution involved. ATHENA is a joint project of a core team of scientists, supported by a broad consortium of knowledge and heritage institutions and public organizations. University partners include Radboud University, University Utrecht, VU University Amsterdam, Wageningen University and the University of Groningen. Coalition partners from the humanities include DANS (data archiving), Nederlab and the Meertens Institute (texts), RKD (iconographic material) and RCE (plant and animal residues). Partners from the natural sciences include SOVON (bird observations), FLORON (plants), RAVON (reptilians, amphibians and fish), Mammal Society (mammals) and Alterra (ecology / vegetation).

The project is subdivided in work packages. For each work package a WP leader is appointed and supported by a steering committee and advisory board. The WP leader, an expert in the relevant scientific field, is responsible for the end result of respective work package. The steering committee is hands-on involved in the work package. The advisory board can provide solicited and unsolicited input on the progress and developments within the work package. Overall project coordination is the responsibility of the program board. The program board yet includes Jan Luiten van Zanden (social economic history), Rob Lenders (historical ecology), Joop Schaminée (vegetation science), Stephan Hennekens (eco-informatics), Theo Spek (landscape history) and Thomas van Goethem (environmental sciences).

## 6. Planning and work packages

### Work packages

The ATHENA project will be developed in a step-wise fashion by means of work packages (WPs). The work packages are subdivided by content and development steps. An overview of the respective work packages is given in table 1, detailed descriptions of the WPs are given below.

**Table 1. Overview of the ATHENA work packages**

<b>WP 0 – Coordination</b>	WP 0.1 Project coordination
<b>WP 1 – Infrastructure</b>	WP 1.1 ATHENA backbone WP 1.2 Metadata WP 1.3 Input portals
<b>WP 2 - Data</b>	WP 2.1 Data model development WP 2.2 Historical data WP 2.3 Natural history thesaurus
<b>WP 3 – Research tools</b>	WP 3.1 Semantic search algorithms WP 3.2 Data visualization
<b>WP 4 - Dissemination</b>	WP 4.1 Demonstrators and symposia
<b>WP 5 - Research</b>	WP 5.1 Research pilots

## **WP0 – Coordination**

### **0.1 – Project coordination**

Objectives:

- Co-ordination of the activities of the project
- Quality assurance of the project's work packages and the processes leading to them
- Overall financial and administrative management of the project
- Management of the project's decision-making processes
- Organization of internal communication

Description of work:

The project coordination will create and support the conditions necessary for a successful and effective collaboration and performance within the project, and coordinate all consortium meetings (plenary meetings, program coordination meetings, steering group meetings, etc.). WP0.1 will liaise with partners and work package leaders to ensure an effective performance monitoring of internal management as well as warranting the best overall performance.

Institutions: UU, RU

## **WP1 – Infrastructure**

### **1.1 – ATHENA backbone**

Objectives:

- Technical inventarisation existing databases
- Linking databases through API's
- Development of the ATHENA interface

Description of work:

WP 1.1 involves the development of the ATHENA infrastructural backbone by linking existing archaeological (BonelInfo and RADAR), ecological (data PGO's, LVD, NLBIF) and historical (Nederlab) databases. First, a technical inventarisation of the participating databases will be conducted to assess the feasibility and requirements for linking the databases. The second step is to link the databases through API's when available. When API's are not available, the existing database needs to be adapted to gain (online) access and successfully link the database to the ATHENA data hub. WP 1.1 will use the results of WP 2.1 'Data model development' to develop the ATHENA interface. At first a basic interface will be constructed, which will be adapted and supplemented with additional functionalities in a later stadium. These functionalities will be developed in conjunction with WP 3 'Research tools'.

Institutions: RU, UU, Alterra, PGO's, RCE, Nederlab

## 1.2 – Metadata

Objectives:

- Explore appropriate formats for metadata
- Publish data as Linked Data
- Apply computational tools for automated metadata creation
- Provide provenance information with the data sets

Description of work:

The various data sources have different characteristics and are therefore not published in a uniform manner. Some data sets are even lacking meta-data or semantic annotations. This makes it difficult to design a uniform format for metadata that is applicable to all species and data types. Computational tools will be used to properly link the different datasets. The data will be published as Linked Data through the use of semantic web technology such as the RDF format, with automatic RDF metadata generation as an option. Furthermore, provenance information, such as information on data transformation, origins of the data and which datasets were used to compute the query, will be made available with the query results.

Institutions: UU, RU, WUR, Alterra

## 1.3 – Input portals

Objectives:

- Inventarisation of the needs of users and database owners
- Explore technical requirements respective databases
- Development input portal infrastructure
- Monitoring functionality and quality input portal

Description of work:

Aim of WP 1.2 is the construction of input portals for the each specific databases/data type. First, the terms and conditions regarding the use and development of data input portals of each institute curating a database will be assessed, as well as the needs of the (end) users of these input portals. Also, the technical requirements for each of the respective databases will be explored. These findings result in data type/database specific protocols for data formats, quality assurance and user interaction. Subsequently the input portal infrastructure will be developed. This will be followed by an extensive period of functionality and quality monitoring. Continuous evaluations will result in further optimizing the process of adding new data to the respective databases.

Institutions: RU, UU, Alterra, PGO's, RCE

## **WP2 – Data**

### **2.1 – Data model development**

Objectives:

- Case studies for selected species
- Inventarisation data model requirements
- Outlook for including larger ecological entities
- Testing preliminary database infrastructure

Description of work:

In WP 2.1 case studies will be used to explore the opportunities, limitations and requirements of developing the ATHENA data model. The case studies for selected species, i.e. various species groups, will focus on availability and heterogeneity of data, data integration and data hub design. The findings of these studies will provide feedback of model requirements and design and give an outlook for including larger ecological entities, such as landscapes. This work package will provide ‘hands-on’ input for other WPs, especially WP 1.1 ‘ATHENA backbone’, WP 3 ‘Research tools’ and WP 5 ‘research’. At the start of the ATHENA project WP 2.1 will give conceptual input to the project WPs. Throughout the project, however, this WP will be used to test the preliminary database infrastructure by means of new case studies. In this process project partners as well as end-users will be involved.

Institutions: UU, RUG, VU, RU, Alterra, (end-users)

### **2.2 – Historical data**

Objective:

- Identifying historical sources
- Digitizing collections (if needed)
- Adding new data to relevant databases
- Creating for databases for (semi-) structured natural history data

Description of work:

Important sources of information are not yet digitized and may contain data that can be geo-referenced, resulting in (semi-) structured data, or, provide important contextual information. To make the information in these sources available an exploratory study will be performed. The resulting inventory holds all necessary information about the available sources and the whereabouts. The next step is to link the (digitized) sources to the relevant existing databases, or, in the case of the (semi-) structured data, include the information in the to be developed historical database.

Institutions: UU, VU

## **2.3 – Thesaurus**

Objectives:

- Development of a thesaurus for selected species
- Exploration automated thesaurus builder techniques
- Expanded thesaurus, for instance landscapes, habitats

Description of work:

A thesaurus is needed to be able to query texts and (semi-) structured data as species are often given common names or synonyms, that generally change through time. Furthermore, in older texts closely related species are often not discerned, species cannot be positively identified (different opinions exist as to which species is actually meant) and 'non existing' species are described. A thesaurus can help with interpreting these sources as it gives insight to the linguistic background. The thesaurus will first be developed for selected species and will include (historical) scientific, Dutch, vernacular and common names, as well as synonyms, etymology and classifications. Larger ecological entities, for instance landscapes and habitats, will probably be included in the project, when this is the case thesaurus can be expanded to include these features. Developing a thesaurus is laborious, so automated thesaurus builder techniques will be explored and applied when considered accurate and feasible.

Institutions: UU, RU, Nederlab, Meertens Institute, PGO's

## **WP3 – Search tools**

### **3.1 – Semantic search algorithms**

Objectives:

- Simultaneously searching digital texts at metadata and text level
- Mining large textual datasets through state of the art computational technology
- Apply tools for contextual search queries

Description of work:

WP 3.1 aims to development/apply semantic search algorithms for queries based on contextual selection. User-friendly tools, for instance developed in Nederlab, allow users to simultaneously search and analyze the digital historical texts made available by scientific libraries and institutions, at text and metadata level. Advances in computational technology can be used for automatic analysis of large quantities of text files for data mining purposes. These tools allow to search and select those text sections from the databases included in the ATHENA data hub that specifically relate to the species of concern, excluding irrelevant (partial) hits based on contextual selection.

Institutions: RU, Nederlab, Meertens Institute



## **3.2 – Data visualization**

Objectives:

- Design of an intuitive user interface
- Adapting and developing state of the art tools for data analysis

Description of work:

The ATHENA data portal will present the data in a clear and attractive way. The shell of the data hub will allow to make sub-selections based on the thematic fields, data sources or fields of research. For data analysis we will draw on tools that have been developed in the individual databases included in the ATHENA data hub. The tools used will be adapted for our specific purpose in presenting the interdisciplinary historical data. State of the art statistical tools will be incorporated that can present, in real time, the query result in a visual fashion.

Institutions: UU, Alterra, Nederlab

## **WP4 – Dissemination**

### **4.1 – Demonstrators and symposia**

Objectives:

- Ensuring efficient communication about the outcome of the project
- Demonstrator meeting
- Promoting interdisciplinary research through scientific symposia

Description of work:

An important part of ATHENA is to increase ‘user involvement’ through so-called demonstrators meeting. These demonstrators show how the available data and tools can be used. By continuously talking to researchers and others users the needs in the area of the digital infrastructure will be identified. Another aim of WP 4.1 is to promote interdisciplinary research on historical human-nature relationships. By organizing symposia around the ATHENA project scientist from the humanities and natural sciences will be brought together.

Institutes: UU, RU

## **WP5 – Research**

### **5.1 – Research pilots**

Objectives:

- Explore the range of tools and features offered by the ATHENA data portal
- Set-up interdisciplinary research lines

- 'Point on the horizon' type research outline

Description of work:

Research pilots provide the opportunity to explore the full range of tools and features offered by the (completed) ATHENA data portal. These pilots can be used to explore new research lines focused on interdisciplinary research, combining data sources and working across scientific domains. These creative processes lead to new ideas and insights. WP 5.1 will provide a research outline for the future of interdisciplinary research on historical human – nature relationships.

Institutions: RU, UU

## Planning

A preliminary overview of the ATHENA project planning is given in table 2.

Table 2. Time planning ATHENA project

WP	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1
<b>0.1 Project coordination</b>												
<b>1.1 ATHENA backbone</b>												
<b>1.2 Metadata</b>												
<b>1.3 Input portal</b>												
<b>2.1 Data model</b>												
<b>2.2 Historical data</b>												
<b>2.3 Thesaurus</b>												
<b>3.1 Search algorithms</b>												
<b>3.2 Visualization tool</b>												
<b>4.1 Dissemination</b>												
<b>5.1 Research</b>												