Research and Innovation Action

Social Sciences & Humanities Open Cloud

Project Number: 823782 Start Date of Project: 01/01/2019 Duration: 40 months

Deliverable 4.20 SSHOCro (final version)

Dissemination Level	PU
Due Date of Deliverable	30/04/2022 (M40)
Actual Submission Date	27/04/2022
Work Package	WP4 Innovations in data production
Task	Task 4.7 Modelling the SSHOC data lifecycle
Type	OTHER
Approval Status	Approved by EC - 27 April 2022
Version	V1.0
Number of Pages	p.1 – p.120

Abstract:

This report describes the SSHOC Reference Ontology (SSHOCro), a common meta-level schema based on CIDOC CRM, to provide a semantic interoperability framework for the description of the data lifecycle used by Social Science and Humanities researchers. The SSHOCro is provided in RDF/S in the file titled "SSHOCro_v.1.0.rdf", which is submitted as an attachment to this report.

SSHOC, "Social Sciences and Humanities Open Cloud", has received funding from the EU Horizon 2020 Research and Innovation Programme (2014-2020); H2020-INFRAEOSC-04-2018, under the agreement No. 823782

History

Version	Date	Reason	Revised by
0.1	26/3/2020	Initial report of deliverable D4.20–SSHOCro class & property definitions - v1.0	Athina Kritsotaki, Eleni Tsouloucha, Chryssoula Bekiari
0.2	21/5/2020	Simplified the labels of properties SHR2, SHR3 to "consists of". Simplified the labels of properties SHR14, SHR15, SHR16 to "follows". Updated the properties' hierarchy list.	Athina Kritsotaki
0.3	25/5/2020	Updated the figures of the schema representation	Athina Kritsotaki
0.4	27/12/2021	Updated the classification of the property "SHR20 has derivative"	Athina Kritsotaki
0.5	04/02/2022	Addition of Terminology, Modelling methodology, Monotonicity–update the text of the introduction	Athina Kritsotaki
0.6	23/03/2022	Redrafted section Modelling Methodology–addition of subsections: 1.5.2.2 Scientific data workflow identified in ARIADNEplus: ancient DNA case-study and CRMhs, 1.5.2.3 SSHOC T5.7 Open linked data; archaeology case study; Redrafted section 1.7 Basic Concepts–addition of subsection 1.7.1 Documenting data using SSHOCro; a simple example Edited figures 5-11 in the usage example (1.8) Edited mappings from Raman spectroscopy in DIAGNOSIS to SSHOCro in the Appendix Overall polishing	Eleni Tsouloucha, Athina Kritsotaki, Chrysoula Bekiari, Maria Theodoridou
0.7	08/04/2022	Peer review	Wolfgang Schmidle (DAI), Holly Wright (University of York)
1.0	20/04/2022	Address peer review comments	Eleni Tsouloucha, Athina Kritsotaki, Chrysoula Bekiari, Maria Theodoridou

Author List

Organisation	Name	Contact Information
FORTH	Chrysoula Bekiari	bekiari@ics.forth.gr
FORTH	Athina Kritsotaki	athinak@ics.forth.gr
FORTH	Eleni Tsouloucha	tsoulouha@ics.forth.gr
FORTH	Maria Theodoridou	maria@ics.forth.gr

Executive Summary

This document serves as the definition of the SSHOC Reference Ontology (SSHOCro, v.1.0). SSHOCro proposes an ontological model and RDF schema to be used as a top-level ontology for organising knowledge and information found distributed across various primary sources of information in the Social Sciences and Humanities Open Cloud (SSHOC).

SSHOCro aspires to provide a semantic interoperability framework for the description of the SSHOC data lifecycle, by offering a conceptual model that can be used to (re)describe at a generic level the real-world lifecycle of creating, finding, and using data–amongst other actions–as it takes place in the various domains of Social Sciences and Humanities. In practical terms, the use of such a model and schema for the research community is twofold: it can be applied as a standard to be used in the step of devising and implementing a metadata capture scheme for tracking the data lifecycle in individual projects, institutions, and disciplines; it can also be used to map, transform and integrate existing data across projects, institutions and disciplines into interoperable pools of information for reuse and exploitation. In this context, keeping track of the processes involved in the data lifecycle amounts to associating each stage with a set of activities performed within it.

SSHOCro is modelled as an extension of the CIDOC Conceptual Reference Model (CIDOC CRM), the standard ontology for Cultural Heritage data, from which it inherits its event-centric orientation and bottom-up approach. CIDOC CRM provides a common and extensible semantic framework that any procedural information can be mapped to. Instances of the CIDOC CRM model can be merged to huge meaningful networks of knowledge about historical facts and contextual relationships (Doerr M., 2003) (ICOM/CIDOC CRM SIG, 2019). The CIDOC CRM model is intended to be a common language for domain experts and implementers to formulate requirements for information systems and to serve as a guide to good practice for conceptual modelling. In this way, it can provide the "semantic glue" needed to mediate between different sources of information, such as that published by museums, libraries and archives.

The development of SSHOCro has further been informed by the output of the mappings of its beta version (v.1.1.3) to well-documented standards used in the Social Sciences and Humanities (SSH)–namely the Data Documentation Initiative (DDI) and the Component Metadata Infrastructure (CMD)I. For more details with respect to the selection criteria and the actual mappings both at schema and instance level, the reader can consult SSHOC Deliverable, D4.19; also (Rasmussen & Blank, 2007; Windhouwer et al, 2012), Since then, a number of mappings with SSHOCro have been undertaken at a schema level, to ensure its the semantic interoperability with the data models describing outputs of the SSHOC Project; and in particular, the data models developed for SSHOC-MP as documented in SSHOC D7.1 Barbot et al.(2019) and SSHOC D7.2 (Ďurčo et al. 2021) and for the SSHOC Conversion Hub (SSHOC format interoperability solutions services) as documented in SSHOC D3.6 (Kleemola et al., 2021). Last but not least, validation of the model has come from the work performed by T5.7 Open Linked Data. Archaeology

Case Study, as reported in SSHOC D5.17 (Schmidle et al., 2020) and in the SSHOC archaeological case study Workshop – The Roman theatre in Catania from survey to interactive 4D visualization, in May 2021.[1] The processes referred to, the sequence in which they take place, the documentation of their input and output information all match the stages of the research workflow referred to by SSHOCro.

The SSHOCro is provided in RDF/S in the SSHOCro_v.1.0.rdf, which is attached to this report. The ontology formalisation in RDF/S (RDF Schema) was created in order to enable the use of the ontology in a wide range of applications accessing registries and knowledge bases.

[1] SSHOC archaeological case study Workshop – The Roman theatre in Catania from survey to interactive 4D visualization, in May 2021. Link to the event description: https://sshopencloud.eu/events/sshoc-archaeological-case-study-workshop-roman-theatre-catania-survey-interactive-4d (accessed March 2022)

Abbreviations and Acronyms

aDNA	Ancient DNA
ARIADNEplus	Advanced Research Infrastructure for Archaeological Dataset Networking in Europe–plus
CERIF	Common European Research Information Format Model
CESSDA	Consortium of European Social Science Data Archives
CIDOC	International Committee for Documentation
CIDOC CRM	CIDOC Conceptual Reference Model
CLARIN	Common Language Resources and Technology Infrastructure
CMDI	Component Metadata Infrastructure
CRMhs	Conceptual Reference Model–Heritage Science
CRMsci	Conceptual Reference Model–Scientific Observation Model
DDI	Data Documentation Initiative
EM	Extended Matrix
EMM	Ethnic and Migrant Minority Survey Registry
EOSC	European Open Science Cloud
ESS	European Social Survey
FORTH	Foundation for Research and Technology–Hellas
FSD	Finnish Social Science Data Archive
HS_	Heritage Science_ (the prefix used for declaring classes and properties in CRMhs)
LCFS	LCFS Living Costs and Food Survey
LINDAT/CLARIAH-CZ	Digital Research Infrastructure for Language Technologies, Arts and Humanities
MP	Marketplace
NLP	Natural Language Processing
ORCID	Open Researcher and Contributor IDentifier
OSF	Open Science Framework
PARTHENOS	Pooling Activities, Resources and Tools for Heritage E-research Networking, Optimization and Synergies
PEM	Parthenos Entities: Research Infrastructure Model
PExx	Parthenos Entity xx (the prefix used for declaring classes in the Parthenos Entities Model; it is followed by a numeric identifier)

PPxx	Parthenos Property xx (the prefix used for declaring properties in the Parthenos Entities Model; it is followed by a numeric identifier)
RDF/S	Resource Description Framework Schema
SEALIT	Seafaring Lives in Transition
SSH	Social Sciences and Humanities
SSHOC	Social Sciences and Humanities Open Cloud
SSHOC-MP	SSH Open Marketplace
SSHOCro	SSHOC Reference Ontology
SO	Scholarly Ontology
SPSS	Statistical Product and Service Solutions
SSRP	Social Sciences Replication Project

Table of Contents

1. Introduction

1.1 Scope

The SSHOC Reference Ontology (SSHOCro) proposes an ontological model and RDF Schema to be used as a top-level ontology for organising knowledge and information found distributed across various primary sources of information in the Social Sciences and Humanities Open Cloud (SSHOC).

It aspires to provide a semantic interoperability framework for the description of the SSHOC data lifecycle, by offering a conceptual model that can be used to describe at a generic level the real-world lifecycle of data produced by the generic workflow of collecting, preparing & connecting, and interpreting data, as well as the auxiliary activities of storing, publishing and finding data–as it actually takes place in the various domains of Social Sciences and Humanities. Its development has, in fact, been informed by data lifecycle management practices in use, in said disciplines. In practical terms, the use of such a model and schema for the research community is twofold: it can be applied as a standard to be used in the step of devising and implementing a metadata capture scheme for tracking the data lifecycle in individual projects, institutions, and disciplines; it can also be used to map, transform and integrate existing data across projects, institutions and disciplines into interoperable pools of information for reuse and exploitation. Within this framework, SSHOCro proposes an ontological model that tries to capture the tools and services used by research communities across the Social Sciences and Humanities disciplines at each point in the data lifecycle, the kind of data they generate/capture, how and by whom the ensuing data are maintained, used, published, and archived, and under what conditions. In that sense, SSHOCro assumes the event-centric approach of digital provenance models, which allows tracing the intermediate results (data) of the processes involved in the research workflows in SSH. In this context, keeping track of the processes involved in the data lifecycle amounts to associating each stage with a set of activities performed within it.

The practical use of SSHOCro lies in publishing and sharing one's research through a data repository. The model can serve as a basis for the mediation of information generated by SSH research, thereby providing the semantic 'glue' needed to transform disparate, localised information sources into a coherent and valuable global resource.

It is designed for resource discovery, i.e., it aims at summarising the most relevant statements about a resource that may help a user locate the information they are looking for within an information system, by exploiting the fact that metadata for the creation, use and discovery of data records are comparable to the information found in the data records themselves.

SSHOCro enables social scientists to document primary data collection, preliminary and interim results, final outcomes, the methods and protocols employed–to name but a few–for all distinct stages of their research. The workflow activities documented through SSHOCro reflect every distinct stage of research

in the SSH and, as such, provide a set of clearly-defined metadata descriptions, conforming to the ontology schema.

Using SSHOCro to document not only one's research outputs, but also the processes by means of which these outputs were generated, allows for their validation and reuse in the context of other research projects. In this sense, SSHOCro can be employed as a high-level workflow logbook.

The proposed ontology has been developed taking the following considerations into account:

1. Cultural and scientific data cannot be understood without communicating the information necessary to understand the data and the ways and circumstances of their creation. This knowledge is comprised of the provenance about the data. It is essential to have metadata created for physical objects as well as for digital objects that bear cultural or scientific interest. Provenance is information about the origin, context, derivation, ownership, or history of some artefact (Doerr & Theodoridou, 2011).

2. Provenance of a resource is a record that describes entities and processes involved in producing and delivering or otherwise influencing that resource. Provenance provides a critical foundation for assessing authenticity, enabling trust, and allowing reproducibility. Provenance assertions are a form of contextual metadata and can themselves become important records with their own provenance (W3C, 2011).

3. Provenance metadata are used to assess meaning (view, experimental setup, instrument settings), relevance (depicted objects, their status, their conditions), quality (calibration, tolerances, errors, artefacts) and possibilities of improvement and reprocessing. They are event-centric and must be described in historical order to ensure the sequence of the events as well as that there are no references to non-existent (non-recorded) events or objects. They encode knowledge about the events that causally relate objects, people, relevant places, and timespans. Alternatively, metadata can document other types of events that have been deduced on the basis of observations (for instance, extrapolations performed on the basis of measurements or other empirical observations) (Doerr & Theodoridou, 2011). In addition, provenance has become even more critical in the web environment where data are sourced not only from established archives, but from many mixed credentialed providers (Lagoze et al., 2013).

4. There are a wide variety of data types and analytical techniques used within and across the disciplines and subdisciplines that constitute the social sciences (Playford et al., 2016).

5. The social science & humanities research can be construed as an iteration of processes like (i) formulating questions (ii) finding empirical evidence (iii) interpretation (inference, causation) (iv) verification in wider context (v) triggering new questions (Doerr et al., 2011)

6. From a data processing point of view, there exists a dominant pattern of activities–collecting data, preparing and connecting data, interpreting the data–in the social science and humanities research, which is supplemented by the following types of auxiliary activity: (i) Persistent Storage, employing physical protected storage or electronic media and curation and access methods (ii) Publication and Presentation, employing, digital file or active database, sites and collections to be visited text, data,

graphics, animation, Virtual Reality, and (iii) Information Selection, employing finding, retrieving, inspecting, and selecting actions.

SSHOCro has been developed in the context of the SSHOC project, the aim of which is to create the social sciences and humanities area of the European Open Science Cloud (EOSC) thereby facilitating access to flexible, scalable research data and related services streamlined to the precise needs of the SSH community.

Relation to CIDOC CRM

SSHOCro uses and extends the CIDOC CRM (Bekiari et al. 2021), a fundamental ontology which aims to link and integrate information relating to cultural heritage, based on a common conceptual approach. The idea at the core of the CIDOC CRM is that [micro-]history can be represented as a series of discrete, largely agreed upon, events, which in turn approximate a part of reality. Events are represented as meetings of people, ideas, objects, which interact with each other in limited areas of space, and time, and bring about noteworthy changes. A particular event can explicitly influence certain actors into a certain course of action, with the effect of indirectly shaping other, subsequent, events. In accordance with this view, SSHOCro addresses the dominant pattern "collection; preparation and connection; interpretation" commonly observed in Social Sciences and Humanities research, along with the auxiliary activities concerning persistent storage, publication and presentation, as well as information selection. More information on the process of identifying this pattern and auxiliary activities can be found in sections 1.5.1 and 1.5.2 of the Introduction. An example showcasing the documenting these core concepts using SSHOCro can be found in the section 1.7 of the Introduction.

The CIDOC CRM aims to provide a common language shared among historians, archaeologists, conservation scientists, museum curators and other cultural heritage specialists to be used in formulating requirements/specifications for information systems; or as a set of best practice guidelines for conceptual modelling; or to be used as a core schema supporting the development of large knowledge networks connected to the Web. Aside from its use in the domain of cultural heritage, the CIDOC CRM has been proved useful in applications in 'electronic' science, (e-science) and biodiversity (Doerr et al. 2014).

The CIDOC CRM standard provides the basic classes and relations devised for the cultural heritage world. The model is complemented by a series of modular extensions. Such extensions are designed to support different types of specialised research questions and documentation needs, such as bibliographic documentation, geoinformatics, archaeological excavations, etc. The extensions of CIDOC CRM are developed in partnership with the research communities in question. These extensions are formulated in a manner that is harmonised with the base ontology: Data expressed in any extension is compatible with the base system of concepts and relations. This harmonised development process leads to a high-level of information integrity and integration not available in other information systems.

The proposed SSHOCro ontology is formulated as a "formal ontology", which is a particular form of knowledge representation. At the same time, it forms an extension of CIDOC CRM, modelling a particular conceptual area that lies outside the scope of the CIDOC CRM; namely the workflows used in data-driven scientific research, oriented to the SSH. The methodology used in the case of SSHOCro expresses the concepts defined in the model as "classes" with binary "properties" linking them together and classifies some of them under existing CIDOC CRM properties and classes.

As such, it represents a specialisation of the core concepts and constructs found in the CIDOC CRM universe. Specifically, it extends CIDOC CRM classes: E7 Activity, E70 Thing, and E55 Type, and relations among classes: P2 has type (is type of),[2] P9 consists of (forms part of), P14 was carried out by (performed), P15 was influenced by (influenced), P16 used specific object (was used for), P33 used specific technique (used by), P130i features are also found on (bears features), P134 was continued by (continued), and P165 incorporates (is incorporated in).

The structure of E7 Activity, following the event modelling, realises an event pattern, which is a strong mechanism for integrating heterogeneous complementary information. An advantage of using the E7 Activity is that all the data elements following this pattern can directly be mapped to the CIDOC CRM. As such, instances of E7 Activity link actions to the actors that committed them, the timespan over which they unfolded, the place(s) where they transpired, the way they unfolded and the changes they brought about in the world (or from which they resulted)–see **Figure 1**, below.

[2] The relations are listed both in their forward going and their inverse form (in parentheses). In the inverse form, the classes linked through the relation are found in a reversed order. For more information, the reader is referred to the "1.3 Terminology" section below. Specifically, entries "inverse of", "domain", and "range".

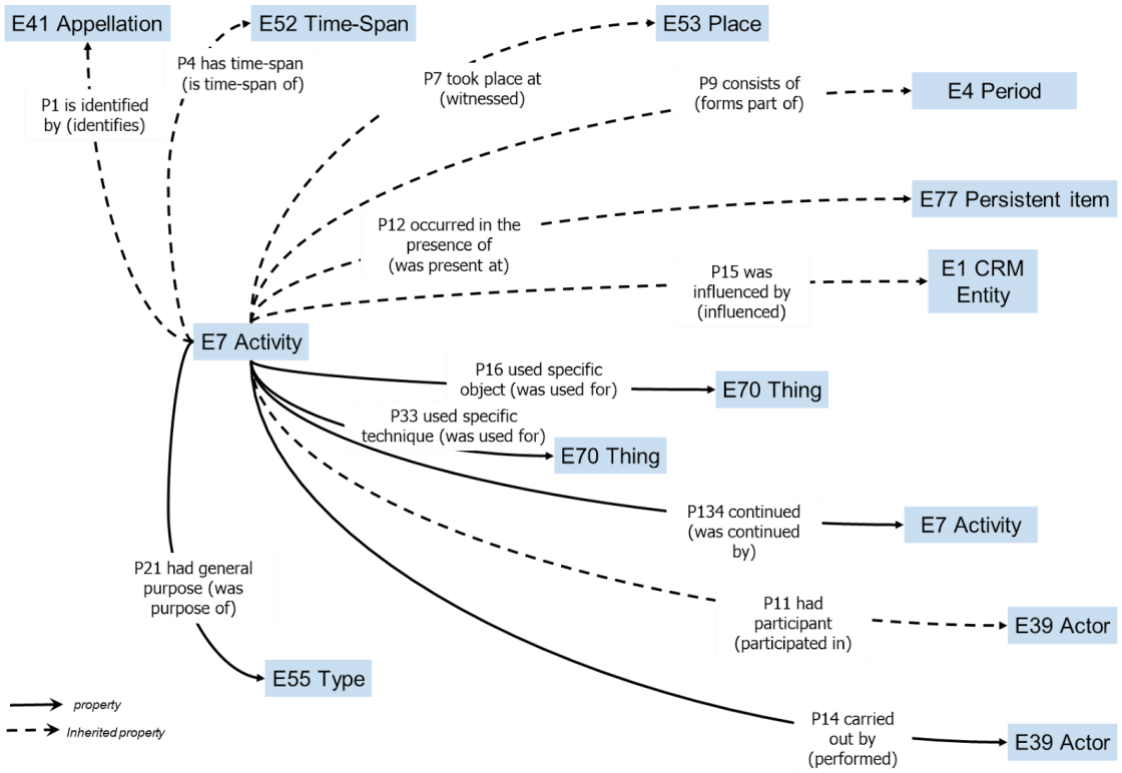

FIGURE 1: CIDOC CRM E7 ACTIVITY

Of the entities represented in **Figure 2**, **Things** correspond to discrete, identifiable, persistent items. They can be material–like any sort of concrete object–or immaterial–f.i. images, texts, datasets, organisational structures etc. Things are involved in activities, in the context of which they are created, used/operated on, modified, or destroyed–depending on the nature of the activity.

Types comprise terms from thesauri and controlled vocabularies used to characterise and classify instances of CIDOC CRM classes. Instances of E55 Types represent concepts (universals).

Places and **Timespans** are used to document the space and time of an event.

Actors (be they individuals or groups) intentionally perform the activities in which they are involved, affecting the things with which they interact in the context of a given activity.

Appellations are used to identify real-world items through names or identifiers.

The transitive relation *P9 consists of* marked on *E7 Activity* is used to designate part-decomposition and structural properties of activities. **Figure 2** above only represents part-decomposition with respect to Activities, as this is considered more relevant for SSHOCro.

Causal relations like influence (f.i., *P15 was influenced by*, *P134 continued*) and motivation (*P21 had general purpose*) for an activity or an event can also be documented with the CIDOC CRM.

Terminology

The following definitions of key terminology provided here, are meant to clarify the precise usage of terms found throughout the present report. Where indicated, they have been adopted from the CIDOC CRM original definition document (Bekiari et al. 2021). Where applicable, an effort has been made to consistently use terminology compatible with the Resource Description Framework (RDF), [3] a recommendation of the World Wide Web Consortium.

Ontology	Ontologies reflect certain views or ways of projecting onto reality; we consider an ontology as a collection of arguments bound by a certain context. The context hereby may be more or less general in its scope and it can be specified (1) according to the kind of ontology at hand, which means according to the view of reality that is adopted; (2) according to the scope of the ontology (e.g., the medical domain vs. the domain of warfare); and also (3) according to the level of granularity at which an inventory of the domain is taken. Factors (2) and (3) ensure selectivity. A list of terms is not an ontology. A data schema is not an ontology.
Formal ontology	A formal ontology is a specification of kinds of things and their relations in terms of logic, approximating reality and can be expressed in a machine-readable form and we can define data structures. They are superior for semantic interoperability and integration.
Class	A class is a category of items that share one or more common traits serving as criteria to identify the items belonging to the class. These properties need not be explicitly formulated in logical terms, but may be described in text (here called a scope note) that refers to a common conceptualisation of domain experts. The sum of these traits is called the *intension* of the class. A class may be the domain or range of none, one or more properties formally defined in a model. The formally defined properties need not be part of the intension of their domains or ranges: such properties are optional. An item that belongs to a class is called an instance of this class. A class is associated with an open set of real-life instances, known as the extension of the class. Here "open" is used in the sense that it is generally beyond our capabilities to know all instances of a class in the world and indeed that the future may bring new instances at any time (Open World). Therefore, a class cannot be defined by enumerating its instances. A class plays a role analogous to a grammatical noun, and can be completely defined without reference to any other construct (unlike properties, which must have an unambiguously defined domain and range). In some contexts, the terms individual class, entity or node are used synonymously with class (Bekiari et al. 2021: 12).

[3] Information about the Resource Description Framework (RDF) can be found at http://www.w3.org/RDF/

Subclass	A subclass is a **class** that is a specialisation of another class (its **superclass**). Specialisation or the IsA relationship means that: 1. all **instances** of the subclass are also instances of its superclass, 2. the **intension** of the subclass extends the intension of its superclass, i.e. its traits are more restrictive than that of its superclass and 3. the subclass inherits the definition of all of the **properties** declared for its superclass without exceptions (**strict inheritance**). In addition a subclass may have none, one or more properties of its own. A subclass can have more than one immediate superclass and consequently inherits the properties of all of its superclasses (**multiple inheritance**). The IsA relationship or specialisation between two or more classes gives rise to a structure known as a class hierarchy. The IsA relationship is transitive and may not be cyclic (Bekiari et al. 2021: 12-3).
Superclass	A superclass is a **class** that is a generalisation of one or more other classes (its **subclasses**), which means that it subsumes all **instances** of its subclasses, and that it can also have additional instances that do not belong to any of its subclasses (Bekiari et al. 2021: 13).
scope note	A scope note is a textual description of the **intension** of a **class** or **property.** Scope notes are not formal modelling constructs, but are provided to help explain the intended meaning and application of the classes and properties. Illustrative example **instances** of classes and properties are also regularly provided in the scope notes for explanatory purposes (Bekiari et al. 2021: 13).
Instance	An instance of a class is a real-world item that fulfils the criteria of the intention of the class. Note, that the number of instances declared for a class in an information system is typically less than the total in the real world. For example, you are an instance of Person, but you are not mentioned in all information systems describing Persons. An instance of a property is a factual relation between an instance of the domain and an instance of the range of the property that matches the criteria of the intension of the property ((Bekiari et al. 2021: 13-4).
property	A property serves to define a relationship of a specific kind between two classes. The property is characterised by an intension, which is conveyed by a scope note. A property plays a role analogous to a grammatical verb, in that it must be defined with reference to both its domain and range, which are analogous to the subject and object in grammar (unlike classes, which can be defined independently). It is arbitrary which class is selected as the domain [...]. In other words, a property can be interpreted in both directions, with two distinct, but related interpretations. Properties may themselves have properties that relate to other classes (This feature is used in order to describe dynamic subtyping of properties). Properties can also

be specialised in the same manner as classes, resulting in IsA relationships between subproperties and their superproperties.

In some contexts, the terms attribute, reference, link, role or slot are used synonymously with property (Bekiari et al. 2021: 14).

inverse of

The inverse of a property is the reinterpretation of a property from range to domain without more general or more specific meaning, similar to the choice between active and passive voice in some languages. In contrast to some knowledge representation languages, such as RDF and OWL, we regard that the inverse of a property is not a property in its own right that needs an explicit declaration of being inverse of another, but an interpretation implicitly existing for any property. The inverse of the inverse of a property is identical to the property itself, i.e., its primary sense of direction (Bekiari et al. 2021: 14).

subproperty

A subproperty is a **property** that is a specialisation of another property (its **superproperty**). Specialisation or IsA relationship means that:

1. all **instances** of the subproperty are also instances of its superproperty,
2. the **intension** of the subproperty extends the intension of the superproperty, i.e., its traits are more restrictive than that of its superproperty,
3. the **domain** of the subproperty is the same as the domain of its superproperty or a **subclass** of that domain,
4. the **range** of the subproperty is the same as the range of its superproperty or a subclass of that range,
5. the subproperty inherits the definition of all of the properties declared for its superproperty without exceptions (**strict inheritance**), in addition to having none, one or more properties of its own.

A subproperty can have more than one immediate superproperty and consequently inherits the properties of all of its superproperties (**multiple inheritance**). The IsA relationship or specialisation between two or more properties gives rise to the structure we call a property hierarchy. The IsA relationship is transitive and may not be cyclic.

In some contexts, the terms object properties are used synonymously with property (Bekiari et al. 2021: 14-5).

superproperty

A superproperty is a **property** that is a generalisation of one or more other properties (its **subproperties**), which means that it subsumes all **instances** of its subproperties, and that it can also have additional instances that do not belong to any of its subproperties. The **intension** of the superproperty is less restrictive than any of its subproperties. The subsumption relationship or generalisation is the inverse of the IsA relationship or specialisation (Bekiari et al. 2021: 15).

domain

The domain is the **class** for which a **property** is formally defined. This means that **instances** of the property are applicable to instances of its domain class. A property must have exactly one domain, although the domain class may always contain

instances for which the property is not instantiated. The domain class is analogous to the grammatical subject of the phrase for which the property is analogous to the verb. It is arbitrary, which class is selected as the domain and which as the **range**, just as the choice between active and passive voice in grammar is arbitrary. Property names in the SSHOCro are designed to be semantically meaningful and grammatically correct when read from domain to range**.** In addition, the inverse property name, normally given in parentheses, is also designed to be semantically meaningful and grammatically correct when read from range to domain (Bekiari et al. 2021: 15).

Range

The range is the **class** that comprises all potential values of a **property.** That means that **instances** of the property can link only to instances of its range class. A property must have exactly one range, although the range class may always contain instances that are not the value of the property. The range class is analogous to the grammatical object of a phrase for which the property is analogous to the verb. It is arbitrary, which class is selected as **domain** and which as range, just as the choice between active and passive voice in grammar is arbitrary. Property names in the SSHOCro are designed to be semantically meaningful and grammatically correct when read from domain to range**.** In addition, the inverse property name, normally given in parentheses, is also designed to be semantically meaningful and grammatically correct when read from range to domain (Bekiari et al. 2021: 15).

Inheritance

Inheritance of **properties** from **superclasses** to **subclasses** means that if an item x is an **instance** of a **class** A, then

1. all properties that must hold for the instances of any of the superclasses of A must also hold for item x, and
2. all optional properties that may hold for the instances of any of the superclasses of A may also hold for item x (Bekiari et al. 2021: 16).

multiple inheritance

Multiple **inheritance** means that a **class** A may have more than one immediate **superclass**. The **extension** of a class with multiple immediate superclasses is a subset of the intersection of all extensions of its superclasses. The **intension** of a class with multiple immediate superclasses extends the intensions of all its superclasses, i.e., its traits are more restrictive than any of its superclasses. If multiple inheritance is used, the resulting "class hierarchy" is a directed graph and not a tree structure. If it is represented as an indented list, there are necessarily repetitions of the same class at different positions in the list (Bekiari et al. 2021: 16).

multiple instantiation

Multiple Instantiation is the term that describes the case that an instance of class A is also regarded as an instance of one or more other classes B1…n at the same time. When multiple instantiation is used, it has the effect that the properties of all these classes become available to describe this instance. In comparison, multiple inheritance describes the case that all instances of a class A are implicitly instances of all superclasses of A, by virtue of the definition of the class A, whereas the combination of classes used for multiple instantiation is a characteristic of

	particular instances only. It is important to note that multiple instantiation is not allowed using combinations of disjoint classes (Bekiari et al. 2021: 16).
monotonicity	The primary role of the model is the meaningful integration of information in an Open World, and so, aims to be monotonic in the sense of Domain Theory. That is, the existing model constructs and the deductions made from them should remain valid and well-formed, even as new constructs are added. A particular consequence of this principle is that no class is declared a complement of a sibling concept under a common direct superclass (Bekiari et al. 2021: 29).
shortcut	A shortcut is a formally defined single property that represents a deduction or join of a data path in the model. The scope notes of all properties characterised as shortcuts describe in words the equivalent deduction. Shortcuts are introduced for the cases where common documentation practice refers only to the deduction rather than to the fully developed path. The model declares shortcuts explicitly as single properties in order to allow the user to describe cases in which he has less detailed knowledge than the full data path would need to be described. For each shortcut, the model contains in its schema the properties of the full data path explaining the shortcut (Bekiari et al. 2021: 16).
Universal	The fundamental ontological distinction between universals and particulars can be informally understood by considering their relationship with instantiation: particulars are entities that have no **instances** in any possible world; universals are entities that do have instances. **Classes** and **properties** (corresponding to predicates in a logical language) are usually considered to be universals (Bekiari et al. 2021: 18).

Naming Convention

SSHOCro classes and properties, linking them to one another, are given both a name and an identifier following the conventions of the CIDOC CRM. The following naming conventions have been applied throughout the proposed SSHOCro reference ontology.

Class labels are identified by numbers preceded by the letters “SHE” (stands for Social science and Humanities Entity–historically, classes were referred to as “Entities”) and are named using noun phrases (nominal groups) using title case (initial capitals). For example, SHE3 SSH Project .

Property labels are identified by numbers preceded by the letters “SHR” (stands for Social science and Humanities Relationship) and. For example, SHR13 follows.

Property names should be read in their non-parenthetical form for the domain-to-range direction, and in their parenthetical form for the range-to-domain direction. Reading a property in range-to-domain direction is equivalent to the inverse of that property. When using a property in reverse direction (inverse of a property), the alphanumeric identifier "**SHRxx**" of the property is followed by “**i**” respectively. For

example, ***SHR3i forms part of*** links from SHE7 Data Interpretation to SHE3 SSH Project , in the inverse order from the forward going property, SHR13 consists of.

Modelling Methodology

The development of SSHOCro follows the conceptual modelling principles listed below:

The creation of the SSHOCro depends on a bottom-up strategy of working from actual, empirical data and corresponding data structures. It emphasises describing the research activities and the scientific processes undertaken by researchers and research institutions, taking into consideration the context in which these activities took place.

What the ontology focuses on describing, is the specific research activities undertaken in the context of SSH, the methods that drive them and their stated goals, as observed on the basis of the data structures documenting them. Literature and practices on a number of topics and scientific domains were consulted (see 1.5.1), but the core of the model was developed on the basis of data structures.

Analysing the research practices and the information requirements of knowledge activities undertaken in the context of research infrastructures was critical for the development of the SSHOCro. To this end, extensive research on SSH data repositories was performed, in order to identify use-cases that would inform the development of the ontology and would be used to validate it. At the same time, SSH data providers were directly contacted for the same purpose.

Note that the goal of developing SSHOCro was not to propose a theory of everything, but rather to model the necessary and well-understood concepts and relations relevant for data-driven research in the Social Sciences and Humanities.

1.1.1 The process of developing SSHOCro

The initial step was to identify workflow patterns used in SSH research, as well as the metadata standards that are often used to document research activities in the scientific domains of interest.

At the same time, an extensive literature review took place in order to assess methods and outputs of the various domains of interest within the SSH universe. It covered topics relevant for Sociolinguistics; NLP; Analytic techniques for Cultural Heritage studies; History (through the examination of workflows

used in history research projects);[4] the study of biodiversity; Statistics and measures for replications of experimental designs in the Social Sciences.

Aside from the literature review, a workshop was organised with the aim of consulting with SSHOC partner-institutions, who are instrumental in the creation or the management and hosting of SSH data. The consultation would allow T4.7 to identify a core set of stages in the research workflow that are commonly observed across scientific domains, as well as target use-cases that would further enrich the SSHOCro and make its content more accessible and better-understood by humans.

During the workshop, it became evident that the research activities documented assume the perspective of the archivist: the concern is primarily about registering the information relevant for the creation and availability of the metadata record, plus the creation and availability of the data or software it reports -- in a very static manner. Reference to methods used for generating and processing data were typically lumped in one section describing the data collection (as an object). The same thing holds for libraries and scripts used for creating a certain piece of software.

In general, associating the methods applied with the kind of end-result they are used to bring about lies outside of the scope of documenting one's research activities. As is registering interim findings that triggered new iterations of data collection or analysis, which also tend to get glossed over, by means of a new release of a dataset or piece of software.

While the ensuing model is an extension of CIDOC CRM, it also incorporates pieces of information coming from both the literature review and the consultation with the community members. At the same time, the development of SSHOCro has been influenced by other ontologies and data models: **CERIF** (Jeffery et al. 2014), **CMDI** (ISO 24622-1, 2015) & (ISO 24622-2,2019), **CRMhs** (Richards et al. 2020), **CRMsci** (Doerr et al. 2018), **DDI** (DDI Alliance, 2014), **PEM** (Bruserker et al. 2018), **SO** (Pertsas & Constantopoulos, 2017), **SSHOC Conversion Hub** (SSHOC D3.6: Kleemola et al. 2021), **SSHOC-Market Place (MP)** (SSHOC D7.2: Ďurčo et al. 2021). Concepts and constructs from said models were considered, however the particular encoding constraints they impose on data structures were not implemented by default.

[4] Workflows used in historic research and their related outputs (in the context of the SeaLiT Project) were consulted for the creation of the model. The task included: a thorough examination of the digitised records of data collected by historians; checking their transcription into electronic data entries; adding metadata upon data entry; corrections by the historians; and, lastly, the interpretation of the data by matching or dispatching instances extracted from the database.
SeaLiT stands for "Seafaring Lives in Transition; Mediterranean Maritime Labour and Shipping and the Challenge of Globalization" and is an ERC Funded Project (ERC ID: 714437). http://sealitproject.eu/?fbclid=IwAR20wxz6iHSscLK4mZGmCRBy-TxK3v0b5RHUBNVACvZixNICWyXlxu4rN8Y

After careful consideration, it was decided that the scenario that best showcased SSHOCro was a series of replication experiments, performed in the context of the *Social Sciences Replication Project* (SSRP), in order to arrive at conclusions about the robustness of experimental data-driven research, and also to assess the predictors of reproducibility for research results (Camerer et al. 2018a, b). In principle, the SSRP formed an ideal example, illustrating how SSHOCro represents research workflow activities: not only were the research protocols, experimental designs and methods thoroughly documented, but the replications involved secondary data collections and reuse of existing data as well.

The development of SSHOCro relied to a large extent on existing documented data, in the sense that its classes and properties formed abstractions from actual data values and their relations.

Abstractions were sought for groups of classes that share a common characteristic–to define high-level generalisations, where necessary, for instance SHE2 Knowledge Workflow Activity.

The practice of defining superclasses for the more specific classes meant that there was no need to define properties that carry identical semantics for as many low-level classes they connect; instead of repeating a property linking a particular tool to a particular type of research activity for all three major types–namely Data Collection, Data Preparation and Connection, Data Interpretation–a property linking from the superclass, Knowledge Workflow Activity to the Tool it made use of was defined: SHE2 Knowledge Workflow Activity. SHR22 used tool (tool was used by): SHE10 Tool. The relation is inherited to the subclasses of SHE2 Knowledge Workflow Activity and can be directly applied to them. This way the model retains its simplicity and economy.

On the other hand, more specific properties or hierarchies of properties have been created, in order to accurately represent the semantics of specific knowledge patterns that follow from actual and relevant research questions for which the information is or can be used.

Upon reviewing existing schemata and data models, an effort was made to distinguish what concepts referred directly to objects or entities and what concepts imply a more complex structure mediated by an event and a set of relations connecting to it. Even when events remained implicit in the data structures examined, the ensuing classes and properties brought them to the fore.

The main concern was to find classes for the relevant properties, and not properties for relevant classes (e.g. Data Interpretation for “interpreted”, Publication for “published”, Data storage for “stored”, etc.). The focus was to detect the general classes for which each property is characteristic. In other words, the model is built by finding the one most specific class that generalises over all classes for which the property applies as domain or range.

Even though classes are deduced on the basis of properties, the definition of said classes should be grounded on the identity conditions of the classes themselves–i.e., as an abstraction over the characteristic properties of the particular instances of the class. Some of the classes are evocative of other models; for instance, Project has also been defined for SO, CERIF, PEM.

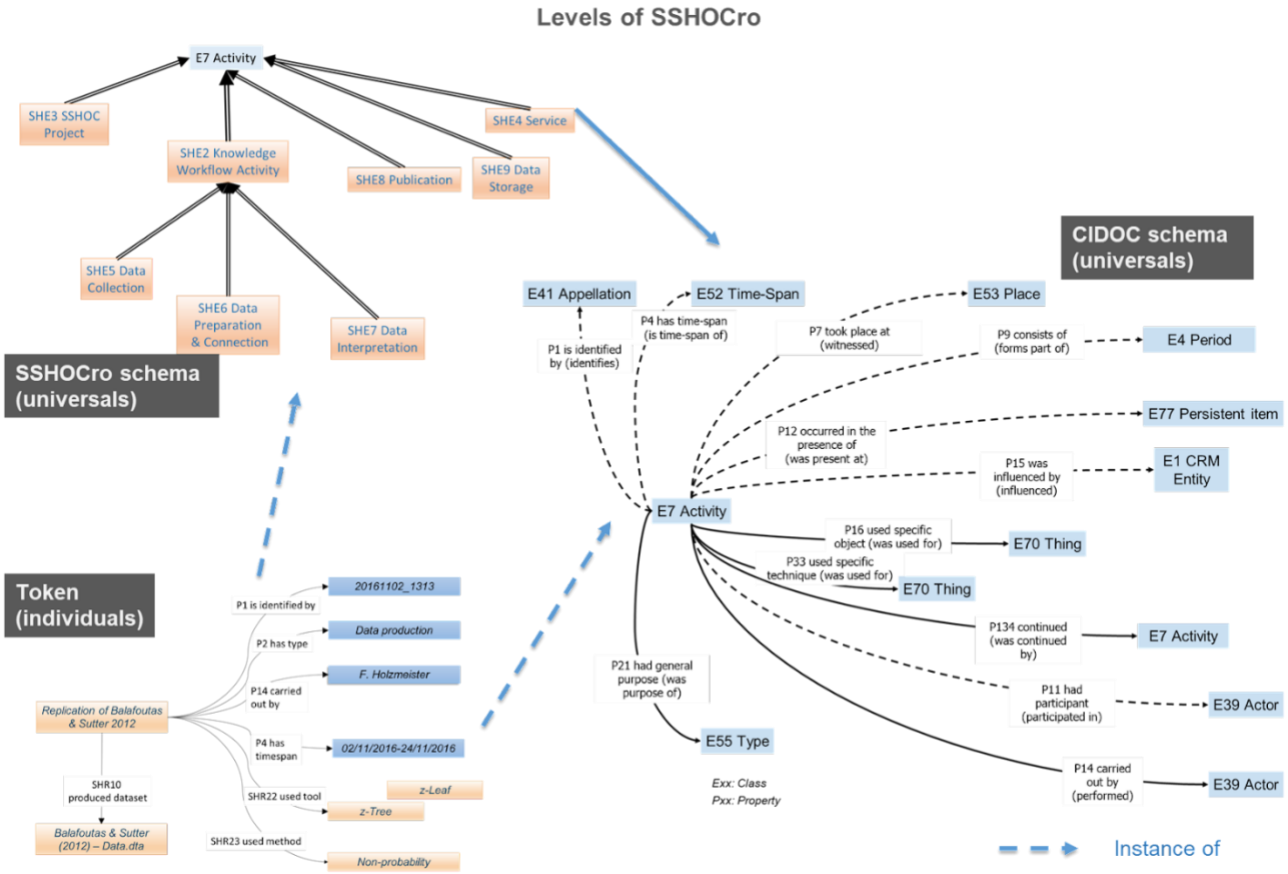

FIGURE 2: GRAPHICAL REPRESENTATION OF THE LEVELS THAT SSHOCRO USES

The methodology applied for the development of the SSHOCro does not apply for creating terminological systems. The design and development of the ontology was an iterative process with several repetitions of the steps described above.

1.1.2 Validation and refinement of SSHOCro

After the initial release of SSHOCro, the ability of the ontology to describe the data lifecycle as it takes place in the various domains of the SSH was subject to testing, through mappings with metadata records expressed in DDI and CMDI. The mappings, described both as a process and as a final output in SSHOC D4.19 (Tsouloucha et al. 2021), served as proof-of-concept and helped identify areas that required further elaborating. The selection of DDI and CMDI as adequate sources for the mappings was motivated by Broeder et al. (2019), according to which said standards were deemed *the two most important ones* for documenting one's research in the SSH. For DDI the records studied were accessed from the FSD, DataverseNo, the EMM Survey Registry, the ICPSR and the SND. For CMDI all examined records came from the LINDAT/CLARIN repository.

A number of mappings at a schema level have been undertaken with a second iteration of SSHOCro as the target schema, to ensure its semantic interoperability with the data models describing outputs of the

SSHOC Project. In particular, the data models developed for SSHOC-MP as documented in SSHOC D7.1 (Barbot et al. 2019) and SSHOC D7.2 (Ďurčo et al. 2021) and for the SSHOC Conversion Hub (SSHOC format interoperability solutions services) as documented in SSHOC D3.6 (Kleemola et al., 2021). Last but not least, the ontology has been presented in a series of workshops, namely the 49th CIDOC CRM SIG meeting (March 11th, 2021), the *SSHOC archaeological case study Workshop – The Roman theatre in Catania from survey to interactive 4D visualization* (May 25th, 2021), and (November 2nd, 2021), where constructive feedback and new use cases were proposed, leading to further refinement of the concepts and properties featured in the ontology.

The decision to test SSHOCro against a broader set of use cases was motivated by the fact that in the metadata records and data models examined by December 2020–when SSHOC D4.19 was submitted–did not explicitly reference a workflow. Instead, the stages of the research documented workflow remained implicit as the records mainly documented the outputs of research activities, with little-to-no indication regarding the activity that generated them. The harmonisation process extended to other data models that purportedly allow one to trace the intermediate results (data) of the research processes they document, i.e., models that associate each stage in the research workflow with a set of activities performed in it.

Use cases examined range from Heritage Science and Archaeology that share the feature of applying analytical techniques–for instance, data from Raman spectroscopy or data for ancient DNA extraction and library creation. Further validation of the model has come from the work performed by T5.7 *Open Linked Data. Archaeology Case Study*, as reported in SSHOC D5.17 (Schmidle et al., 2020). Of these, the data model deployed for **DIAGNOSIS: Cultural Heritage Laser Analysis and Diagnosis Documentation System**–a documentation system that supports advanced imaging, analysis and diagnosis of Cultural Heritage objects, collaboratively developed by the Institute of Electronic Structure and Laser (IESL-FORTH) and the Institute of Computer Science (ICS-FORTH) in the context of the project POLITEIA[5]–was used to implement mappings to SSHOCro.

1.5.2.1. Mapping the fields describing the application of Raman spectroscopy in Diagnosis to SSHOCro

The "DIAGNOSIS" system supports the long-term preservation of the data produced during scientific examinations of analytical techniques performed by the "Photonics for Heritage Science" group of IESL-FORTH. The system exploits XML technology, multilayered architectures, open-source software, and

[5] POLITEIA-I: "Advanced analytical, diagnostic, surveying and documentation technologies in Cultural Heritage", Action for the Strategic Development on the Research and Technological Sector, (NSRF 2007-2013), MIS-448300, Coordinator: D. Anglos (http://www.forth.gr/politeia/)

international standards, and provides remarkable capabilities concerning data organisation, documentation process and system architecture and system functions.

Core entities of the data model of "DIAGNOSIS" are (i) **scientific examination**, which documents all the data produced by the application of a scientific examination, (ii) the **object**, which documents the object under consideration, (ii) the **device,** which describes the infrastructure used by a specific scientific examination.

Scientific examinations vary with respect to the methods they apply, and the procedures they follow in collecting and processing information, but they follow a similar workflow, which has been mapped to SSHOCro. This way the workflow patterns observed for Raman spectroscopy that are reflected in the system of "DIAGNOSIS", can be used to verify the scientific research workflow patterns defined in SSHOCro. The main stages assumed in SSHOCro–namely, Data Collection, Data Preparation and Connection, and Data Interpretation–have all been identified in the workflow of "DIAGNOSIS". Furthermore, the tripartite pattern assumed by SSHOCro is corroborated by Romary et al. (2019)–i.e., the research scenario describing the analytical protocol followed for Raman spectroscopy analysis directly on a cultural heritage object, where no sampling is required. The methods and techniques applied are directly linked to the research activities associated with a particular stage in the scientific workflow described. The mapping to SSHOCro had as input the XML data exported by the system DIAGNOSIS. The XML data exports document a scenario such as the one described by Romary et a. (2019). The output of the mapping consists of the re-expression of the XML nodes used to document the techniques and activities in a Raman analysis, into the equivalent statements in SSHOCro.

The mappings and information on the case study can be found in the appendix.

1.5.2.2. SCIENTIFIC DATA WORKFLOW IDENTIFIED IN ARIADNE$_{PLUS}$: ANCIENT DNA CASE-STUDY AND CRMHS

A typical scenario of scientific analyses applied to archaeological investigation (Richards et al., 2020) include the following steps:

1. Start by collecting objects and samples from their original archaeological site or context. Document all the information relevant to the archaeological context in question, as well as all the environmental conditions and the collection procedures.
2. Transfer the collected objects and/or samples to the lab, where measurements and further preparations take place.
3. Analyse samples using well-defined research protocols and instrumentation. Intermediate reports documenting results are generated for every sub-task in the analysis phase.
4. Integration and/or revision of the results of the analysis to enhance their precision and truthfulness.
5. Dissemination of the results (datasets and analyses).

The application of such scenarios can be extended to Heritage Science as a whole. Of the steps identified above, the objects documented in situ (No.1) serve as input for the Data Collection (SHE5) that generates the original dataset, to be augmented in the next stages of the research. In a similar vein, documenting the actions undertaken to prepare the objects and samples for analysis in the lab, as well as documenting the types of analyses they are intended for (No.2) correspond to Data Preparation and Connection (SHE6). Documenting the analyses–in terms of methods, materials, instrumentation–and the observations they yielded (No.3), as well as the integration and revision of the results obtained (No.4) correspond to Data Interpretation (SHE7), as defined in SSHOCro.[6]

These research activities are thoroughly documented; during each of them are generated datasets, where the objects in question and their properties are completely represented.

The digitisation process through which the physical objects or samples and the observations and measurements that concern them lies outside the scope of SSHOCro, which, in its turn, captures the knowledge generation process in data driven research, at a high level of abstraction. It is implicit in the figure below, where the processes of collecting data and preparing them to be interpreted are informed by the things–objects or samples–that the data represents.

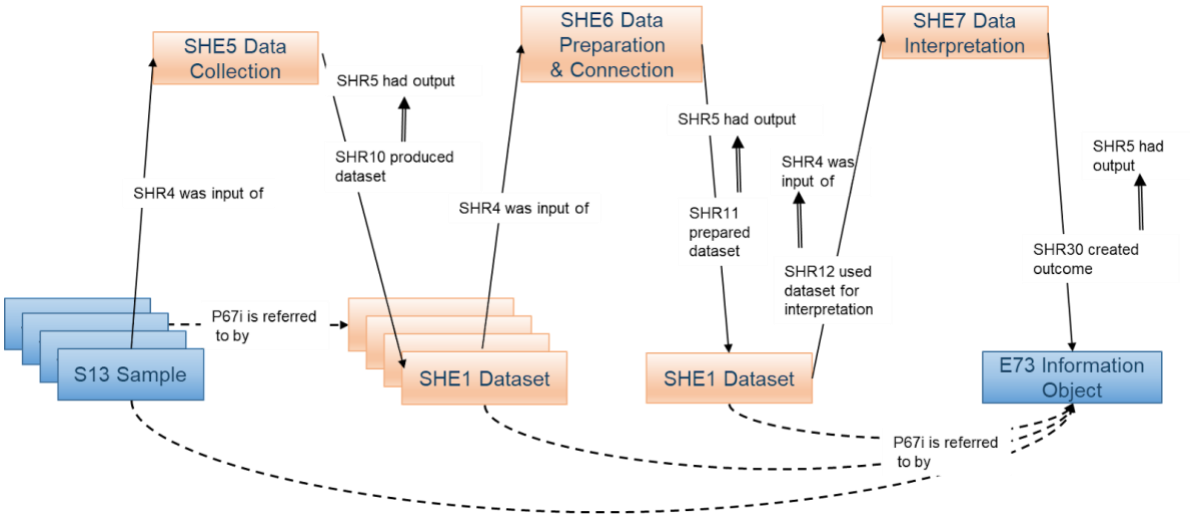

FIGURE 3: AN ITERATION OF DATA COLLECTION, PREPARATION & CONNECTION, AND INTERPRETATION FOR A-DNA

A specialisation of the above workflow applies to the definition of a model describing ancient DNA (aDNA) wet lab services, offering further support to the model proposed by SSHOCro. An analysis of the projects that were undertaken by the aDNA lab of the Institute of Molecular Biology and Biotechnology (IMBB) at

[6] Despite the fact that SSHOCro foresees the possibility of disseminating research outcomes at all stages in the workflow (SHE8 Publication can be the output of all the stages of a scientific research workflow), in this particular instance, the authors refer to hypotheses and interpretations that evaluate the outcomes of the research (step No.5)

FORTH yielded the following steps–for which correspondences with the SSHOCro can easily be established at an abstract level: (1) maps to SHE5 Data Collection, (2-3) map to SHE6 Data Preparation and Connection, and (4) maps to SHE7 Data Interpretation. Publication activities (5) are, again, associated with the publication of the overall results in the form of a research paper and supporting data.

1. Sampling
2. DNA extraction & Library construction for aDNA; a stepwise procedure that consisted of the following subtasks:
    1. Drilling of teeth/petrous bone
    2. DNA extraction
    3. Adapter preparation, building library
    4. QPCR Quantification
    5. Indexing PCR
3. Sequencing and bioinformatics analyses
4. Interpretation of the results and conclusions
5. Publication of results

The stages of the workflow identified were aligned with CRMhs, an application ontology developed to capture and represent concepts relevant for Heritage Science.[7] It results from a harmonisation of different existing ontologies in use for modelling research activities and provides entities and constructs particular to Heritage Science. The model supports the documentation of the archaeological excavation process and covers wet lab research activities that are highly relevant for several scientific domains, such as Bioarchaeology and ancient DNA analysis, Environmental Archaeology, Archaeology of inorganic materials, Dating of archaeological findings, etc. Where available, mappings to CRMhs are provided for SSHOCro classes and properties.

1.5.2.3. OPEN LINKED DATA; ARCHAEOLOGY CASE STUDY

Further validation of the model has come from the work performed by T5.7 *Open Linked Data. Archaeology Case Study*, which presents an example of an actual transition of archaeological excavation data to cloud,

[7] The CRMhs model is composed of a set of independent classes used for distinguishing and defining each of the entities involved in a specific scientific analysis, and of a set of relationships used for linking these entities between each other, according to the specific sequences of events in which they are involved.
"Study Object", one of the main entities of the model, is defined as any physical object or artefact that is subject to scientific investigation and analysis, such as archaeological finds, human remains and other organic and inorganic materials; art works. Typical scenarios documented by CRMhs include the process of collecting objects and samples from their original archaeological site as well as performing measurements and scientific observations on said objects.
The development of the model has been informed by the outputs of the project ARIADNEplus.

as reported in SSHOC D5.17 (Schmidle et al., 2020). One of the issues the partners of SSHOC T5.7 have decided to address is the transparency of virtual archaeology projects. Documenting all action points undertaken in the context of a 3D reconstruction allows validating the results of the reconstruction without having to duplicate the effort from scratch.

The scientific research workflow identified by SSHOC T5.7 documents the following steps:

1. Documentation: the literature review, which in this case extends to both written and iconographic sources related to the object of investigation, as well as structured data–such as archaeological excavation data.
2. Survey: the collection of graphic documentation data using techniques like laser-scanning, digital photogrammetry, photo and video shooting of the area of interest. Instrumentation is also specified.
3. Data Processing: establishing connections among collected data points, f.i. enhancing 3D models of in situ elements with 3D models of non-in situ objects, implementing stratigraphic readings using Harris Matrix relations (through the EM tool),[8] etc.
4. 3D-Reconstruction Hypothesis: optimization of 3D models, texturing, illuminations etc.; validation of the reconstructed models.

Basic Concepts

The following paragraphs explain SSHOCro basic concepts with the help of graphical representations. SSHOCro is built around the core notion of **CIDOC CRM E7 Activity**.

The relations linking instances of E7 Activity to the entities necessary for describing them are inherited by the activities specifically defined for SSHOCro.

A more detailed schema is developed under this core structure, as can be seen in the following figures. **Figure 4**, represents the temporal entities of SSHOCro–all subclasses of E7 Activity; **figure 5**, represents the endurants particular to SSHOCro that are involved in the said activities, plus their relation to CIDOC CRM E77 Persistent Item and (relevant) set of subclasses.

[8] Extended Matrix; a tool for virtual reconstruction in Archaeology: http://osiris.itabc.cnr.it/extendedmatrix/

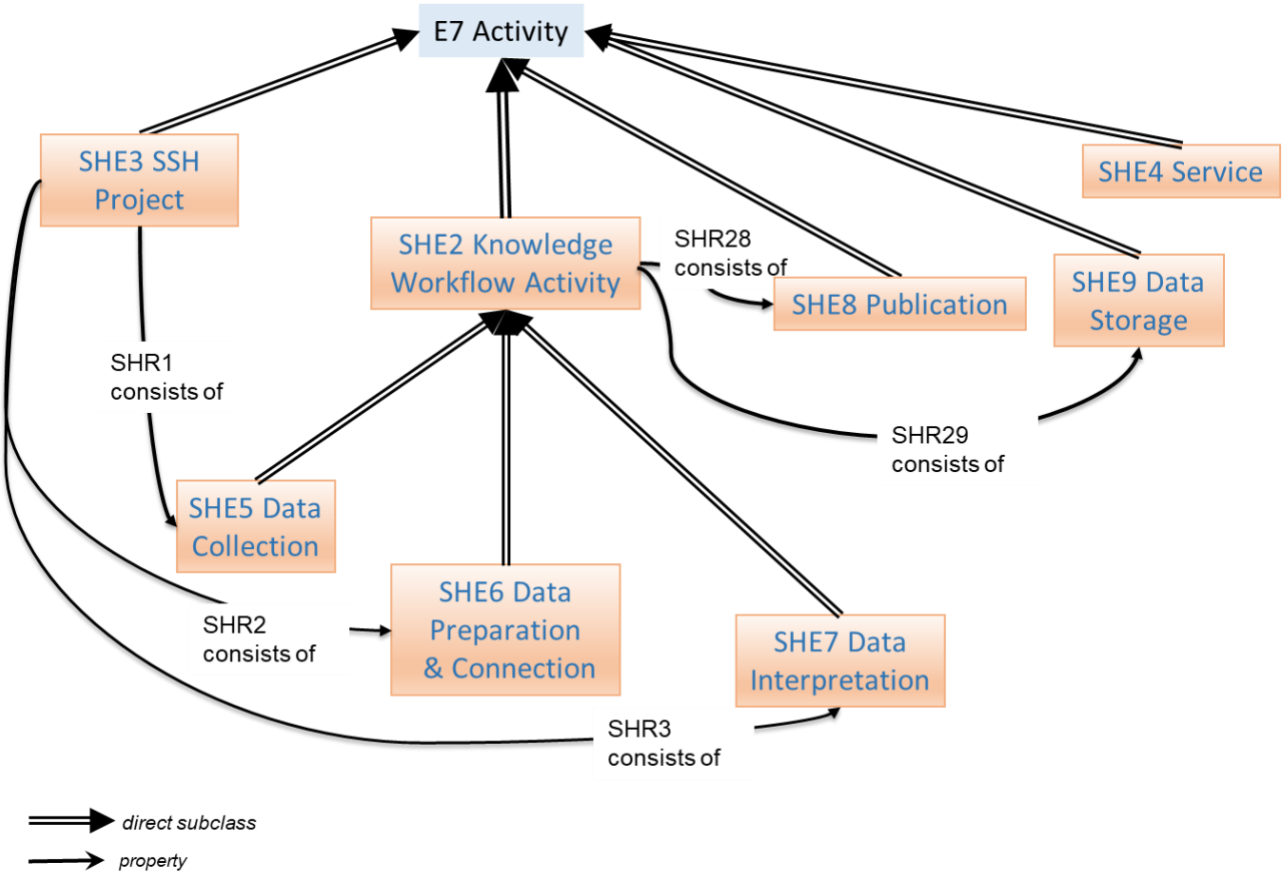

FIGURE 4: SSHOCRO TEMPORAL ENTITIES

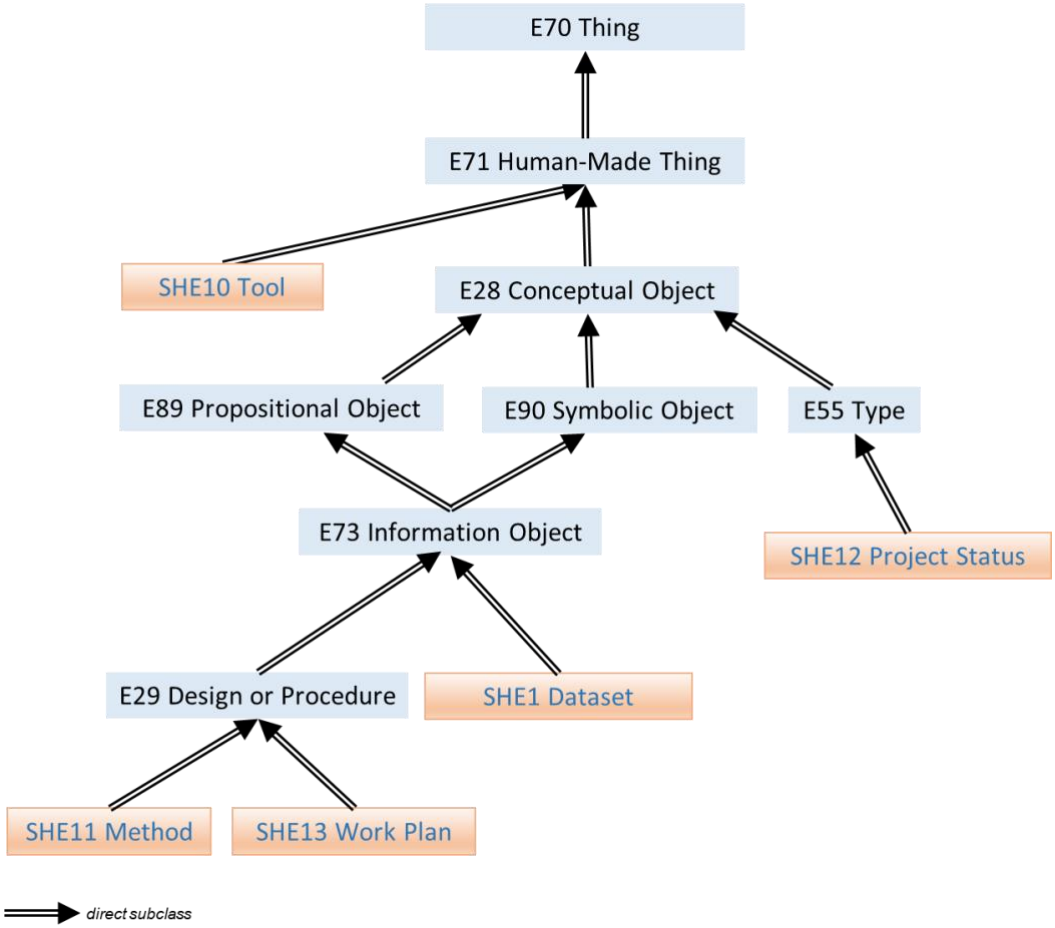

FIGURE 5: SSHOCRO THINGS

Based on empirical research and especially on the analysis of the working practices of researcher communities, the proposed model describes the processes identified in, and documented by, most scientific research workflow procedures. To this end it makes use of a small set of concepts representing basic and distinct phases of knowledge production and research activities.

The knowledge workflow describing said scientific procedures is a generic activity, which involves physical and information objects as inputs and outputs. The stages identified in the context of this activity are the **data collection** phase -i.e., the processes involved in collecting data (qualitative and quantitative); data collection is usually followed by a distinct phase comprising the **preparation and connection of datasets** -e.g., how to treat missing values and outliers, and/or the process of annotating a text by identifying the names of persons or places, for instance, across the datasets generated in the data collection phase and subsequent data preprocessing and other manipulations. These activities can be continued by the process of **interpreting datasets**, by means of examining or comparing data in order to test theories and/or offer a plausible explanation regarding the examined phenomena. The model treats the documentation of the series of activities undertaken at each stage as an instance of publication.

The three activity types that can form part of an SSH Project are specialisations of SHE2 Knowledge Workflow Activity, which refers to the actions and processes executed in the course of scientific research (see **Figure 6**). These properties range from relations to (i) activities that concern publishing and storing data and other outputs of the research; (ii) activities that form the basis for deploying new iterations of the processes in question; (iii) inputs and outputs generated during a phase of a research workflow activity; (iv) tools, services and methods associated with stages in the workflow.

For instance, the relation with SHE10 Tool can refer to objects and software used for data collection (1), tools and software used for data connection and preparation (2), instrumentation and software used to analyse the data in the data interpretation stage.

1. **SHE5 Data Collection:** Baskets, mattocks, shovels, brushes are all tools one would find in an archaeological excavation, where the objective is the acquisition of objects and samples. Hard copies of questionnaires are tools used in surveys to collect data values for the variables of interest. Pieces of software are used to automatically register respondents' responses to different treatments in an experimental setup or in a web-survey.
2. **SHE6 Data Preparation & Connection**: Dilutants and other substances used in preparing samples in a wet-lab environment, software used to document samples and the conditions with which they were obtained would all be documented as tools that one would use in a wet-lab setting during data entry. In a social survey setting, the software used to replace missing data points with values calculated according to an imputation technique or creating dummy variables where necessary would be examples of tools relevant for the data preparation and connection stage.
3. **SHE7 Data Interpretation:** The software used for analysing the data to derive the statistical measures that allow one to make inferences regarding their scientific hypotheses.

On the other hand, properties that contribute to identifying the distinct phases of a research workflow activity that define an instance of SHE3 SSH Project (*consists of*) and represent all iterations of the knowledge generation process within it (*follows*), were only defined for the specialisations of SHE2.

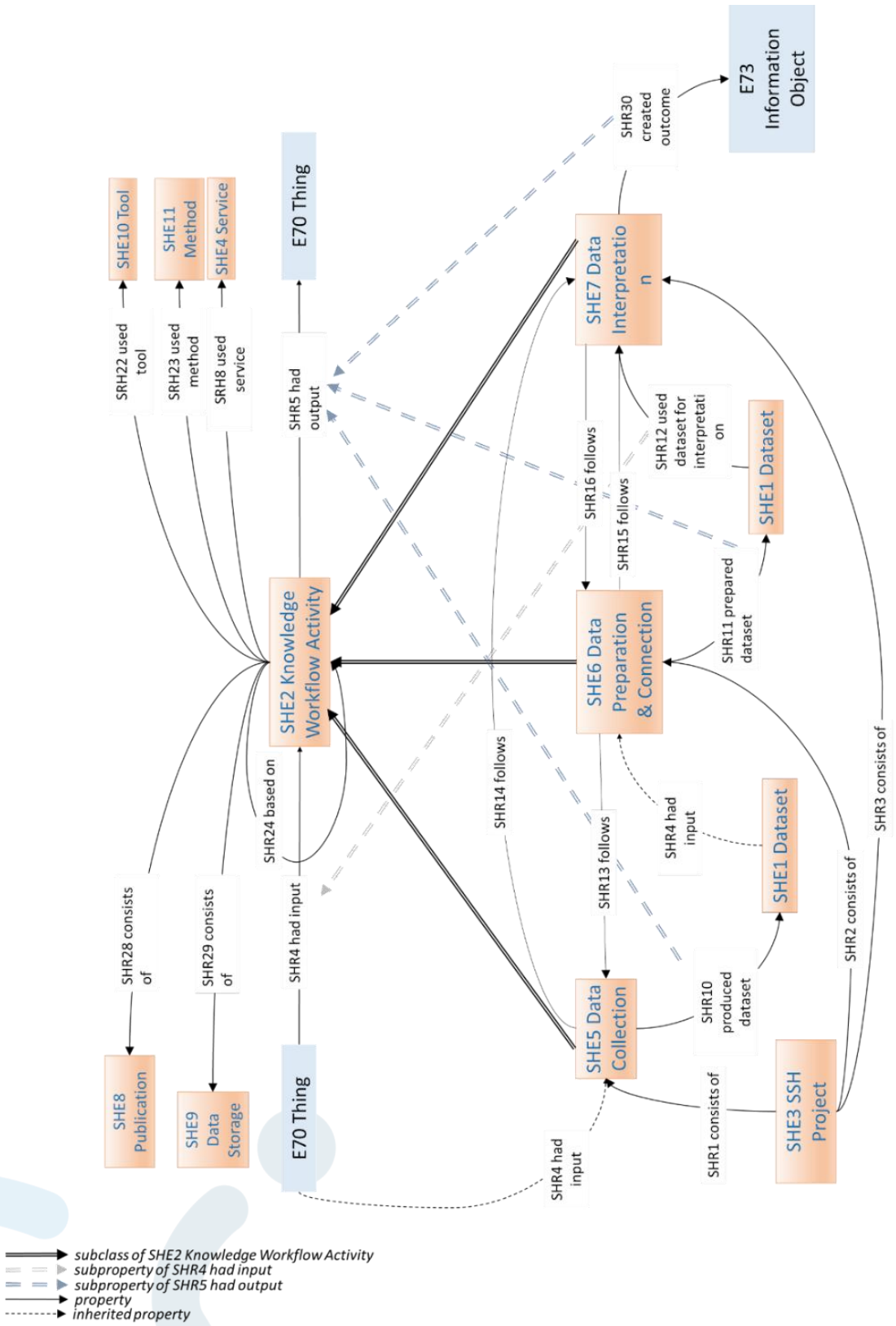

FIGURE 6: SSHOCRO SHE2 KNOWLEDGE WORKFLOW ACTIVITY

The workflow proposed here is not a sequence of procedures, one following the other in a linear and predetermined order. Instead, it captures an open process, whereby one can always backtrack to alter bits and pieces of the procedures followed, in an iterative manner, which reflects the stages involved in the scientific process -i.e., how knowledge is produced and verified. Defining the workflow activities relevant for an instance of SHE3 SSH Project at a more specific level than SHE2 ensures that no information concerning the research activities executed in one particular iteration of the workflow, nor their relation to other activities in subsequent iterations, can be lost.

Figure 7 demonstrates possible sequences of knowledge workflow activities, as these have been identified from literature review, data structures, and consultation with data producers:

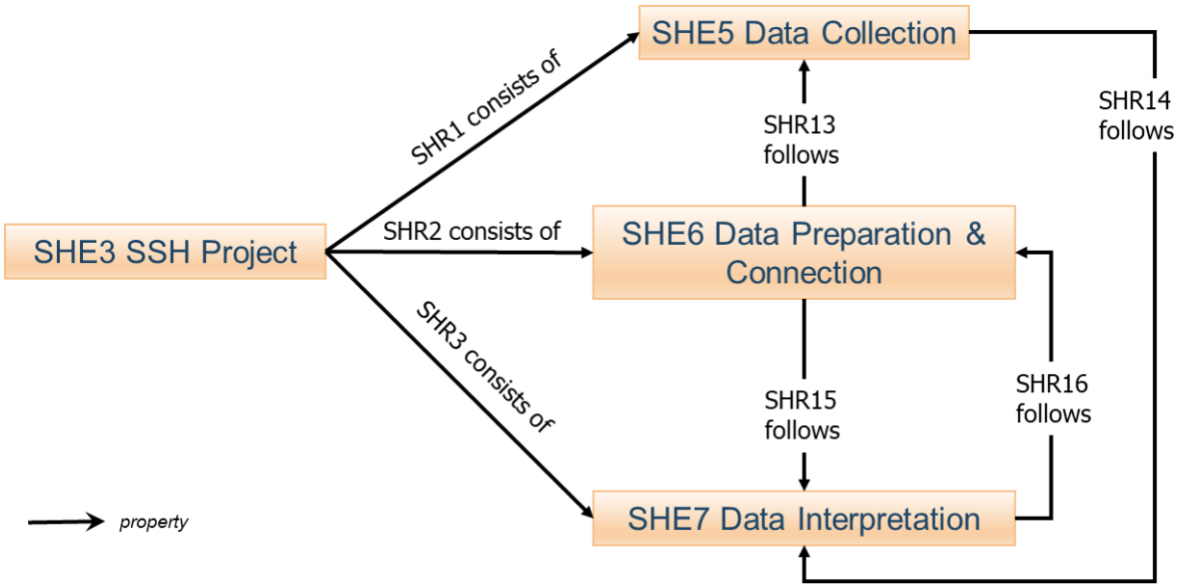

FIGURE 7: PARTHOOD RELATIONS AND TEMPORAL RELATIONS AMONG SSHOCRO ISA SHE2 ACTIVITIES

1. Arrows pointing from an earlier to a later stage in the research workflow (SHR13 follows, SHR16 follows) represent a "canonical" sequence of research activities: researchers collect the data they need for their research, they prepare them/connect them, and they interpret them. The research knowledge cycle has been completed.
2. On the other hand, arrows pointing from a later to an earlier stage in the research workflow (SHR14 follows, SHR15 follows) indicate that a backtracking process has occurred. Examples for that include:
    a. Upon failing to reject the null hypothesis (SHE7 Interpretation for SSHOCro) for a dataset that indicates a strong correlation between a number of variables, the researchers resort to a second data collection (SHE5 Data Collection), to increase their sample size and rerun the statistics.
    b. During a data exploitation and interpretation phase, in which theories have been developed for recognising instances of persons, locations, artefacts etc. that–despite referring to the same real-world entity–they have been documented as distinct, it is very

often is the case that curating the datasets generated in the context of these activities involves researchers revisiting the transcriptions of the documents in question and correct the entity names or properties (in that case, SHE6 Data Preparation follows SHE7 Interpretation)–see example for SHR15 follows (is followed by).

c) Declaring a "follows" relationship for SHE2 Knowledge Workflow Activity, in both domain and range position, to be inherited by its subclasses would obscure the succession of the discrete stages and the notion of workflow altogether. The set of activities that are particular to each stage in the workflow and their relations to other such activities would become implicit.

d) The process is "open" in the sense that it need not be the case that an instance of SHE5 Data Collection was followed by an instance of SHE6 Data Preparation and Connection stage, or that the latter was followed by an instance of SHE7 Data Interpretation.
    a. For instance, upon documenting the activities performed in the context of a project of manuscript digitisation, it is possible that the digitised outputs be published as soon as one has obtained the digital copies. The activities documented would fall under "collecting data".
    b. In a similar vein, if one is interested in producing a machine searchable digital copy of said manuscripts, they would have to implement a series of activities corresponding to the extra step (SHE6 Data Preparation and Connection). The machine searchable digital outputs could be published straight away.
    c. However, if one wanted to proceed with analysing the topics of the said digitised manuscripts, they would not need to perform more operations on the data (i.e., they would engage in SHE7 Data Interpretation activities). The same thing would apply if they were interested in implementing a critical edition of the manuscripts.

Section 1.6.1 and 1.6.2 form usage examples for the SSHOCro.

1.1.3 Documenting data using SSHOCro; a simple example

Assuming that someone needs to document the dataset–a digital file, in essence–entitled "daF3062e.csv". This dataset is the output of the survey Finnish Attitudes to Immigration, for which the data collection had the form of individual subjects' responses to a self-administered questionnaire. It is documented that the data was collected on the 12th and 13th of August 2015 in Finland.

To document how the data collection phase resulted in the production of the dataset, one needs to link SSHOCro class SHE5 Data Collection to the generated dataset. Users need to consult the definition document of SSHOCro in order to determine the appropriate relation through which to declare the link. In this case, the appropriate relation seems to be SHR10 produced dataset (dataset was produced by), by means of which one connects instances of SHE5 Data Collections to instances of SHE1 Dataset that they generated.

In order to declare this connection, the user should first assign identifiers to the process and its output. Once this has been implemented, only then can they document the dataset "daF3062e.csv" in their local/public triple store repository using the following triples.[9]

1st step: identify the following instances of SSHOCro classes:

- "**S1**" serves as an identifier representing **"Taloustutkimus survey, Finnish Attitudes to Immigration (2015)"**, and instantiates SSHOCro class SHE3 SSH Project
- **"PH1"** serves as an identifier representing the **"Collection of the data values (individual participants' responses to a self-administered questionnaire)"** to be examined in the context of the survey, and instantiates SSHOCro class **SHE5 Data Collection**
- **"D1"** serves as an identifier representing the **dataset "daF3062e.csv"** generated by the data collection, and instantiates SSHOCro class **SHE1 Dataset**

This is expressed through the following statements. Note that the statements do not constitute an instance of implementing the ontological model proposed by SSHOCro into an actual information system. To achieve that, one must take into consideration the technical environment that the ontology will be translated into. The technical aspects do not lie in the scope of the present document, and will only be addressed insofar as there is an expressed interest by research institutions to document research activities using SSHOCro.

- "Taloustutkimus survey, Finnish Attitudes to Immigration (2015)". *P1 is identified by (identifies)*:[10] **S1**
- Collecting the individual participants' responses to a self-administered questionnaire. *P1 is identified by (identifies)*: **PH1**
- "daF3062e.csv". *P1 is identified by (identifies)*: **D1**

2nd step: start to add more meaningful statements, by using the specific relationships of SSHOCro f.i.,

- D1. *dataset was produced by*: PH1 (SHE5)–the dataset is declared a product of the particular phase in the research workflow

[9] Where necessary, they can use CIDOC CRM classes and properties.
[10] *P1 is identified by (identifies)* is inherited by CIDOC CRM

Given that T4.7 is willing to maintain the ontology–i.e., to offer updates, demonstrate known applications, collect feedback and provide guidelines for implementations and further deployments–a designated website has been created that can be accessed through this link: https://isl.ics.forth.gr/SSHOCro.

1.1.4 Class and property usage examples

The aim of this section is to illustrate by means of an example how SSHOCro classes and properties are used and linked to one another. The example is drawn from material published through the Open Science Framework (OSF) in the context of the Social Sciences Replication Project (SSRP).[11] In 2016, a group of researchers undertook the task of replicating 21 experimental studies published in *Nature* and *Science* magazines between 2010 and 2015, to determine the robustness of the reported results and the validity of the replication methods in general. These replications were conducted in the context of the Social Sciences Replication Project,[12] initiated by the Center for Open Science.[13]

The goal of the SSRP was to replicate experimental studies with as minimal deviations as possible from the original experimental designs, procedures and methods followed in all the stages of the original studies. Replicating experimental studies allows one to assess the robustness of results arrived at through experimental observation. Given that successful replications increase the trust in scientific findings, they are considered pivotal, especially when they involve influential studies that predate the development of new methods, approaches and/or data. Multiple research groups were engaged in the Social Sciences Replication Project, each assigned to replicate a set of experiments, which, in their turn, included one or two distinct stages of ***(i) data collection, (ii) data preparation and connection, and (iii) data interpretation.*** Whether one or two rounds of replications took place depended on the success or failure of the experiment to replicate on the first round. The analysis stopped after the second round, irrespective of successfully replicating the results obtained by the original experiment.

The study chosen to exemplify the classes and properties of SSHOCro, was the replication of *Balafoutas & Sutter (2012)*, the hypothesis of which was that preferential treatment increases women's competitiveness, without reducing the competitiveness of men, in a lab environment. The results extrapolate in policymaking.

[11] For more information on the project–in terms of overall goals, methods used, institutions and research teams involved, evaluations etc.–the reader is referred to Camerer C.F. et al. (2018). *Evaluating the replicability of social science experiments in Nature and Science between 2010 and 2015*. Open Science Framework. Retrieved February 25, 2020, from https://osf.io/pfdyw/

[12] All materials for the Social Sciences Replication Project (2016) are made available via the corresponding Open Science Framework (OSF) directory: https://cos.io/our-services/research/ssrp-overview/

[13] The designated website of the Center for Open Science is: https://cos.io/

The researchers who performed the replicating study (F. Holzmeister, J. Huber, M. Kirchler, J. Rose) followed the experimental design of the original study, used the same tools for data collection, analysed them using the same tests and packages and arrived at, essentially, the same conclusions. The data, the procedure and the software for collecting it, the packages and the scripts used for data analysis are available for the replication through the OSF.[14] However, they are not available for the original study.

In what follows, the reader can navigate through diagrams that document the information relayed by the example.[15]

Figure 8 below represents the *Replication of Balafoutas and Sutter (2012)* at a macro-level. The *Replication of Balafoutas and Sutter (2012)* instantiates a Project, which forms part of a larger Project, undertaken by the group of researchers from the Innsbruck University division of the SSRP community. The SSRP, in its turn, is the overall Project in the context of which the aforesaid Projects form a part of. The SSRP as a whole *offered a service*, namely the creation of a repository for replication studies in the Social Sciences, comprising materials (i.e., datasets plus scripts and software to operate on them) as well as reports and/or other publications. This transitive parthood relation is represented in the diagram by *the P9i forms part of* relations among the different sub-projects and the overall one. The instance of SHE3 SSH Project, *SSRP_Replication of Balafoutas and Sutter (2012)*, can be further broken down to the series of discrete sets of activities, each corresponding to a different stage in the research (collection, preparation and connection, and interpretation).

The goal of the Project of replicating the experiment by Balafoutas & Sutter (2012) was to test the accuracy of predictions made by the authors in the original version of the experiment. This implies a relation of influence between the replication and the original study, captured through *SHE3 SSH Project. P15 was influenced by: SHE3 SSH Project*.

The rest of the properties connecting the *SSRP_Replication of Balafoutas and Sutter (2012)* are inherited by *E7 Activity*–a superclass for *SHE3 SSH Project*.

[14] Holzmeister F., Huber J., Kirchler M., Rose J. 2018. *Replication of Balafoutas and Sutter (2012)*. Open Science Framework. Retrieved February 25, 2020, from https://osf.io/m8qav/
[15] Cf. Figure 1: CIDOC CRM E7 Activity.

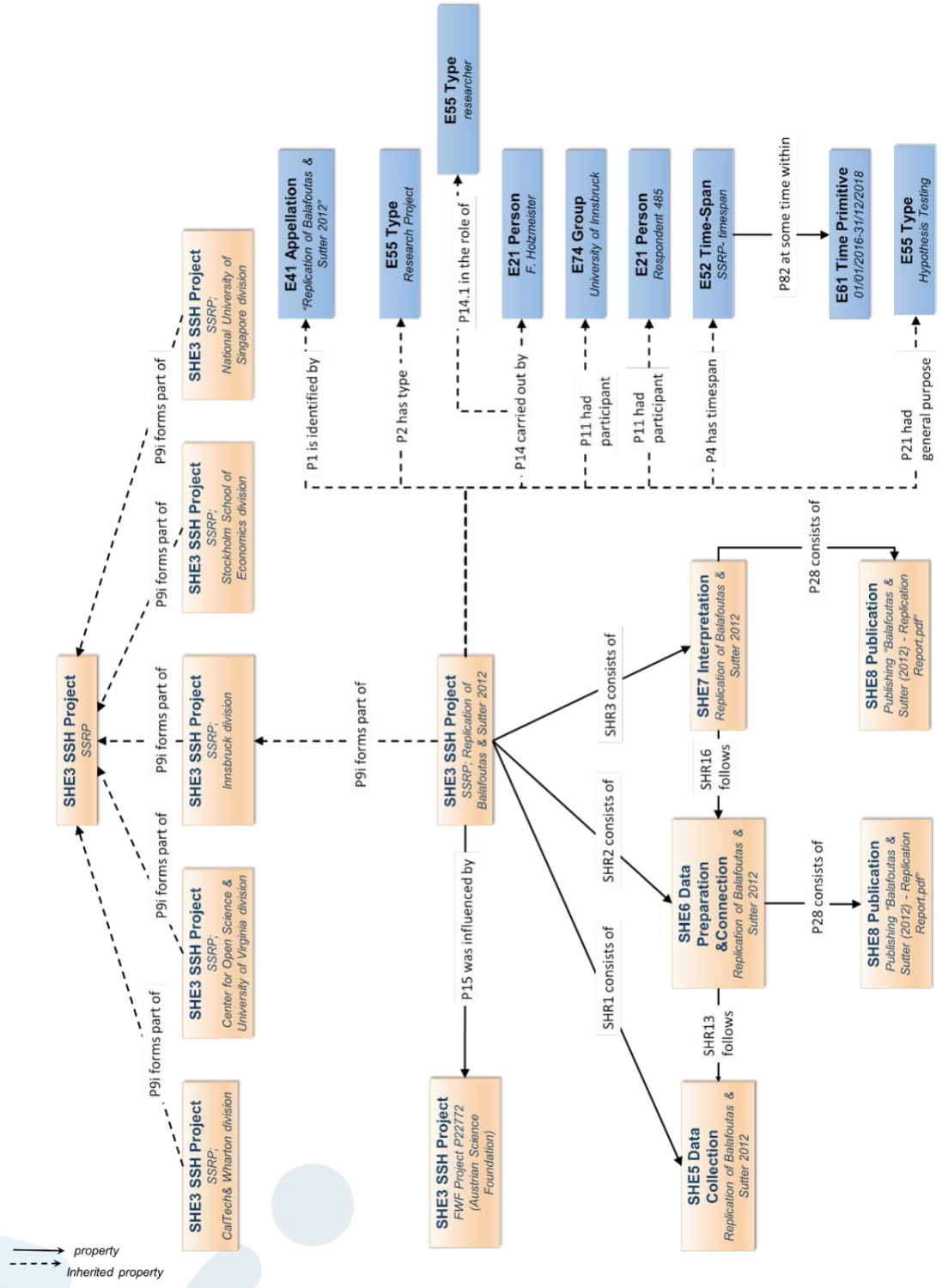

FIGURE 8: REPLICATION OF BALAFOUTAS & SUTTER (2012) AT A MACRO-LEVEL

Replication studies depend to a great extent on the scientific methods used in the studies, the results of which are tested. That dependence is manifold and relies upon as little deviation as possible from the research protocols observed in the original studies, on behalf of the teams conducting the replication.

The processes and methods undertaken during data collection, data preparation and connection, and data interpretation in the replicate study are bound by the ones observed in the original. Replications are data driven and it is important that the relation among any data produced and manipulated in the context of the replication and the original study be truly comparable. In that sense, one must be able to express how they relate to one another, which is what **Figure 9**, below, aims for–namely to express the relationship of datasets to the projects that produced or deployed them in the course of a research project, and the relationship between datasets across different projects.

In this particular case, the Dataset "Balafoutas and Sutter (2012) – Analysis.do" that contains the scripts for the statistical analyses undertaken in the original experiment, forms part of the datasets used by the original experiment. It is reused in the replication of the experiment by Holzmeister and colleagues in the context of SSRP. It was used as input for the activities performed in the SHE7 Data Interpretation stage in the research workflow.[16] The unshaded part of the diagram represents the data reuse at a Project level: namely, it showcases how "Balafoutas and Sutter (2012) – Analysis.do" serves as input for SHE7 Data Interpretation. The shaded part shows how the dataset in question is linked to other datasets generated in the course of this particular Project (it is incorporated in them) and which stages of the overall research workflow it is relevant for.

[16] The dataset containing the R-scripts that was used during the analysis was obtained during the data collection stage of the research workflow. This connection is not portrayed in Figure 2, as it would overburden it.

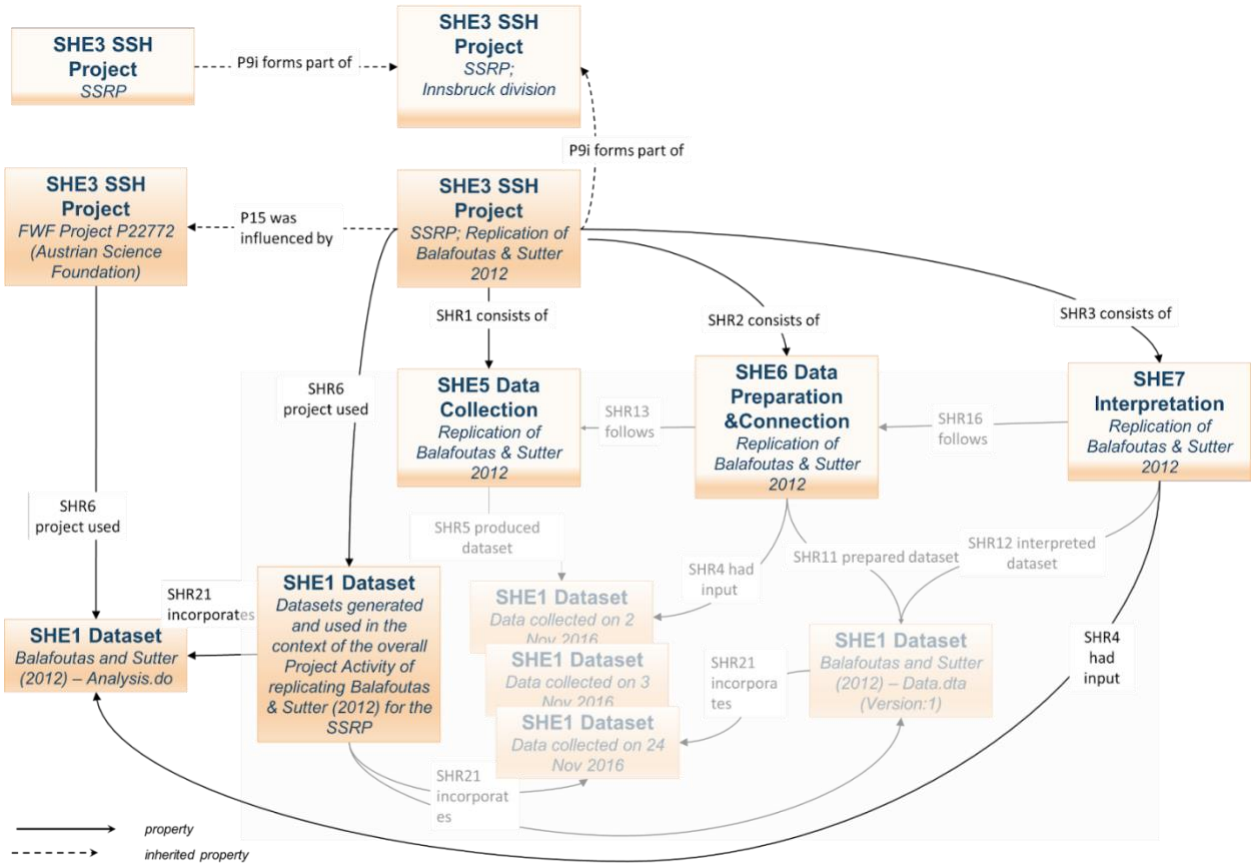

FIGURE 9: DATA REUSE IN SSRP SUBPROJECTS

Figures 10 through 13 illustrate the sets of activities undertaken during each of the distinct phases of the Project *SSRP_Replication of Balafoutas and Sutter (2012)*. In particular, **Figure 10** illustrates the data collection process, **Figure 11** the stage of data preparation and connection, **Figure 12** the process of interpreting the collected data (post cleaning), and **Figure 13** the events of publishing the results or the data itself. Publishing activities can take place at any stage during the research project–we have opted to demonstrate it for the interpretation stage, because it's most expected during that stage.

Figure 10 represents the data collection stage; it comprises the overall collecting of data and links it to the datasets generated in its course, any information on the actors that participated in it (either the researching team, who was responsible for running the experiment or the institution that provided the infrastructure necessary for its running), the timespan over which it occurred (which is inferred by the dates of the distinct data collection rounds), the software used for collecting the data (z-Tree and a client application for subjects, z-Leaf), and the method used for the sampling procedure by means of which the subjects that participated in the experiment were selected. According to the replication report (Holzmeister et al., 2021a), the overall sample was pooled from students attending Innsbruck University. Their placement into separate treatment groups was random, however, since the results are used to draw conclusions over the entire population (vs. male and female students at Innsbruck University), the method is characterised as "Non-probability".

The overall purpose the data collection serves is arbitrarily taken to be "method verification", in accordance to the attested goals of the project–i.e., testing the validity of predictors used for replications in Social Sciences experiments.

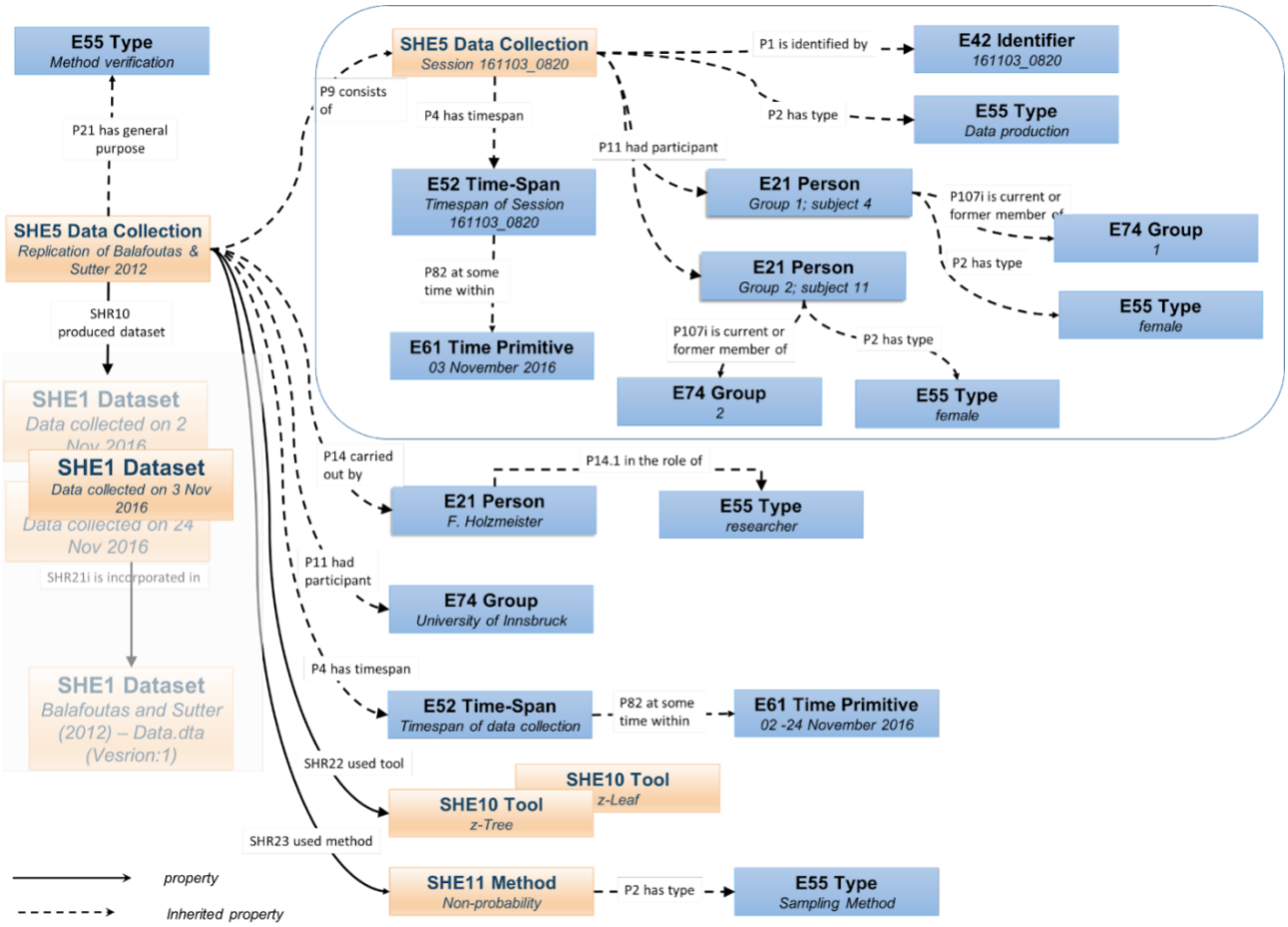

FIGURE 10: THE DATA COLLECTION PROCESS; COLLECTING INDIVIDUAL OBSERVATIONS VS. GENERATING DATASETS THAT CONTAIN THEM

As demonstrated in the circumscribed section of **Figure 10**, the data collection phase can be subdivided into distinct actions of collecting individual observations: like for instance the actions of collecting six observations per individual, per group, for Groups 1 through 3–each consisting of 6 subjects (3 male, and 3 female).[17] The type of the data collection can refer to whether data was actually produced or found to be reused. It can be applied wherever there is relevant documentation.

[17] The representation in the figure glosses over the particulars of collecting observations per group, per individual and per treatment, but this is done on purpose–to increase its legibility.

Figure 11 represents the data preparation and connection stage; it documents the overall preparation and connection of the data that have been produced earlier on in the project (during the data collection phase). The datasets created by each data collection process–at the level of the observations of the individual subjects' responses to the various treatments, per group–were incorporated in the dataset "Balafoutas & Sutter (2012) – Data.dta (Version:1)" that was the output of the data preparation and connection phase.

Aside from listing information on the actors that performed the activity and the time when the relevant actions took place,[18] **Figure 11** also links the stage of data preparation and connection to the software of statistical analysis that the data were entered into (Stata).

The authors report a miscalculation of the sample size required, which resulted in a deviation of 6 observations (Holzmeister et al., 2021a). Even though no reference is made to the method applied for the missing values treatment, in Figure 10 it assumed that it involved some kind of imputation method.

The publication node refers to the activity of releasing the dataset, "Balafoutas & Sutter (2012) – Data.dta (Version:1)", following data entry in Stata. The time of the publication is taken to be the date it became available through OSF.

[18] The timespan of the data preparation and connection is not documented, but can be inferred to have occurred between the last day of the data collection round–namely, 24th November 2016–and the date that the dataset "Balafoutas & Sutter (2012)–Data.dta (Version:1)" appeared on OSF–namely, 20th December 2017.

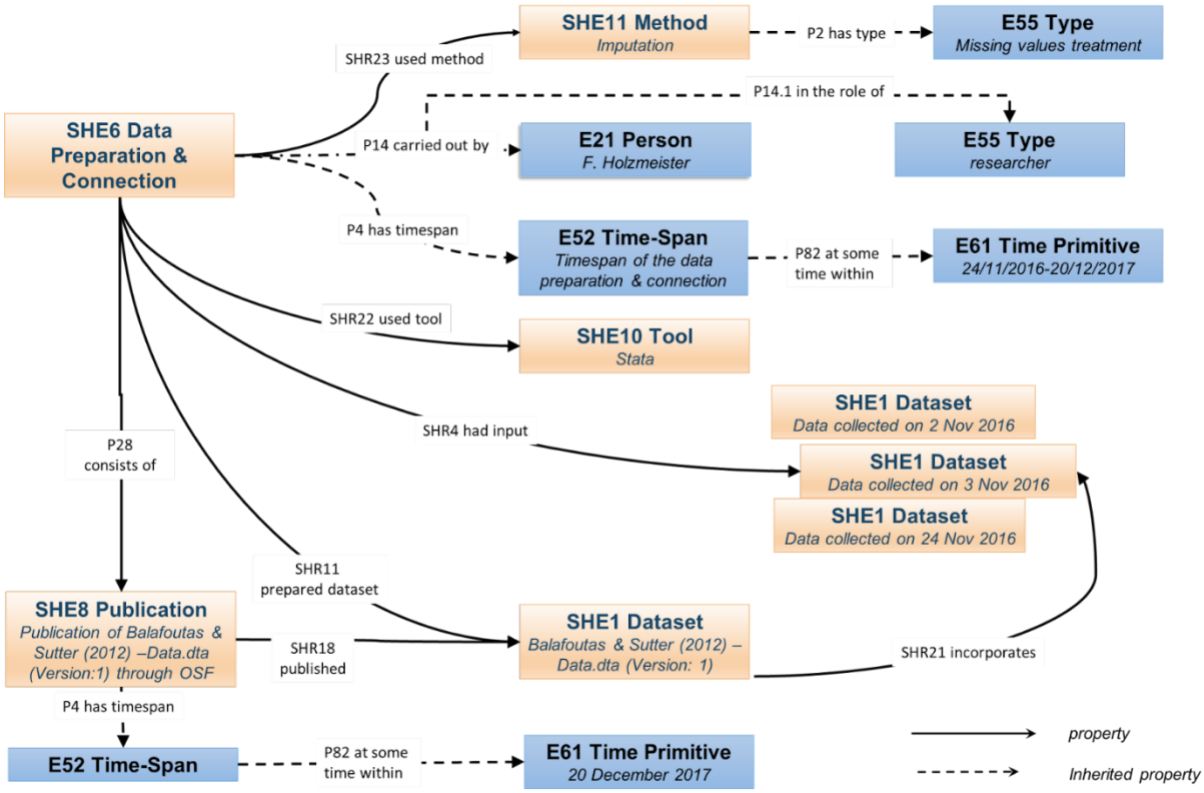

FIGURE 11: THE DATA PREPARATION AND CONNECTION PROCESS (PRE-PROCESSING PHASE)

Figure 12 represents the data interpretation stage in the research–namely all actions relevant for the process of validating or disproving hypotheses based on data analysis; the actions undertaken in this stage involve operations on the previously collected and preprocessed dataset named "Balafoutas & Sutter (2012) – Data.dta (Version:1)"–documented through *SHR12 interpreted dataset:SHE1 Dataset*.

The scripts listed in the dataset documented through *SHR4 had input:SHE1 Dataset* are the ones obtained for reuse through the data collection process, namely *"Balafoutas and Sutter (2012) – Analysis.do"*. The operations listed within (calculation of statistical measures) serve as an interpretation of the dataset, as they allow the researchers to test the validity of the hypotheses put forth, and correspond to the methods for calculating said measures.

The relation between the data analyses and the software used to perform them is also documented through the property SHR22 used tool.

Figure 12 also links the activity of interpreting the results of the experiment with the person or group that performed the analysis, and with the timespan over which it occurred. Last but not least, it offers information on the auxiliary activity of publishing one's results (SHE8) in the form of a replication report. Publishing activities are represented in greater detail in the following diagram.

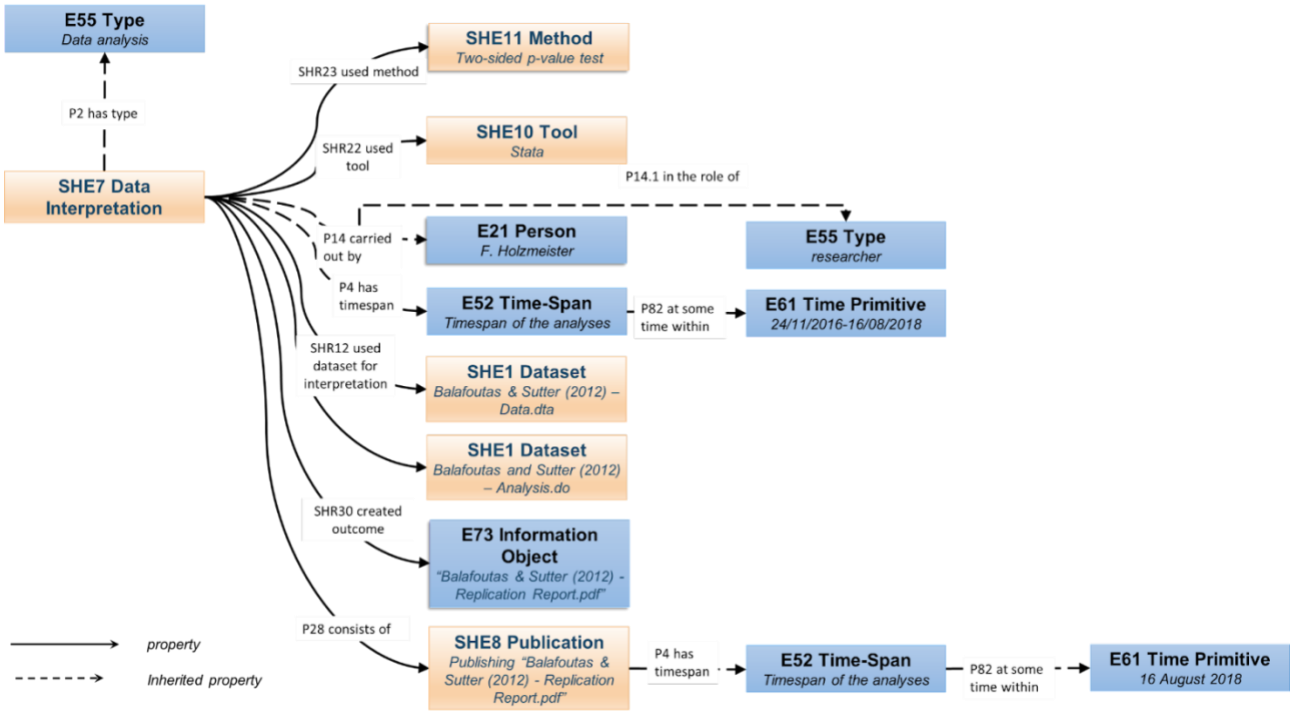

FIGURE 12: THE DATA INTERPRETATION PHASE

Aside the three knowledge generation processes that form specifications of SHE2 Knowledge Workflow Activity, there is a set of auxiliary processes that can form part of it, namely publication, which amounts to the activity of disseminating the outputs of a knowledge generation process and making them publicly available–taking into considerations any conditions or restrictions on their use; and data storage, which amounts to the activities of creating back-ups for the data and other outcomes of knowledge generation process according to policies and practices observed by the interested research communities. Publication and data storage form auxiliary research activities, insofar that they share or store the outputs of a research workflow stage, but they do not define distinct stages of a research workflow.

Figure 13 illustrates the publishing of the materials and the reports for the replication of Balafoutas and Sutter (2012) through the OSF repository. It links the publication activity with the stage of the research workflow it forms a part of, the output that it published, the actor who performed it, the purpose of publishing it (disseminate the results), and the time it occurred.

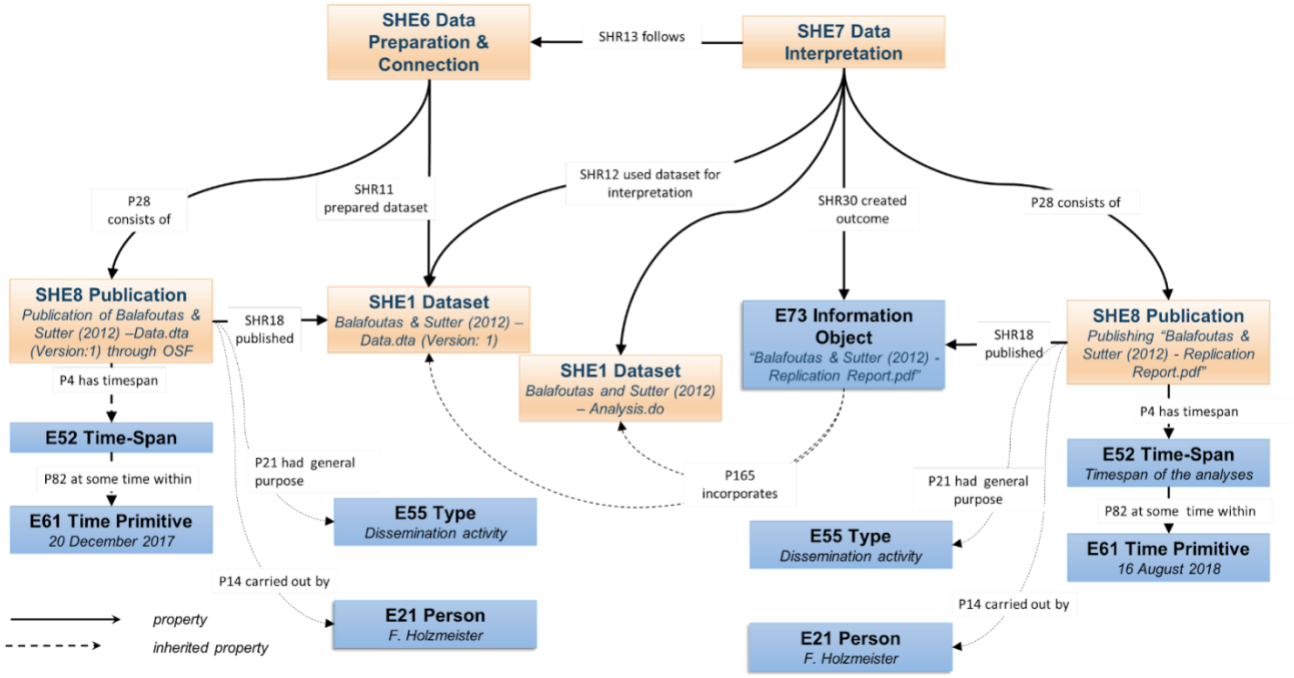

FIGURE 13: PUBLICATION OF DATA AND RESULTS

2. SSHOCro Class hierarchy aligned with (part of) the CIDOC CRM Class Hierarchy

— E1 CRM Entity
— — E7 Activity
— — — SHE3 SSH Project
— — — SHE4 Service
— — — SHE2 Knowledge Workflow Activity
— — — — SHE5 Data Collection
— — — — SHE6 Data Preparation and Connection
— — — — SHE7 Data Interpretation
— — — SHE8 Publication
— — — SHE9 Data Storage
— — E70 Thing
— — — E71 Human-Made Thing
— — — — SHE10 Tool
— — — — E28 Conceptual Object
— — — — — E90 Symbolic Object
— — — — — — E73 Information Object
— — — — — — — E29 Design or Procedure
— — — — — — — — SHE13 Work Plan
— — — — — — — — SHE11 Method
— — — — — — — SHE1 Dataset
— — — — — E89 Propositional Object
— — — — — — *E73 Information Object*
— — — — — — — *E29 Design or Procedure*
— — — — — — — — *SHE11 Method*

— — — — — — — — *SHE13 Work Plan*

— — — — — — — *SHE1 Dataset*

— — — — — E55 Type

— — — — — — SHE12 Project Status

3. SSHOCro Property hierarchy aligned with (part of) the CIDOC CRM Property Hierarchy

PROPERTY NAME	ENTITY-DOMAIN	ENTITY RANGE
P14 carried out by	E7 Activity	E39 Actor
— SHR25 provided by	SHE4 Service	E39 Actor
P9 consists of (forms part of)	E4 Period	E4 Period
— SHR1 consists of (forms part of)	SHE3 SSH Project	SHE5 Data Collection
— SHR2 consists of (forms part of)	SHE3 SSH Project	SHE6 Data Preparation and Connection
— SHR3 consists of (forms part of)	SHE3 SSH Project	SHE7 Data Interpretation
— SHR28 consists of (forms part of):	SHE2 Knowledge Workflow Activity	SHE8 Publication
— SHR29 consists of (forms part of):	SHE2 Knowledge Workflow Activity	SHE9 Data Storage
P2 has type	E1 CRM Entity	E55 Type
— SHR9 has status	SHE3 SSH Project	SHE12 Project Status
P15 was influenced by (influenced)	E7 Activity	CRM Entity
— SHR24 based on (was basis for)	SHE2 Knowledge Workflow Activity	SHE2 Knowledge Workflow Activity
— P16 used specific object (was used for)	E7 Activity	E70 Thing
— — P33 used specific technique (was used by)	E7 Activity	E29 Design or Procedure

PROPERTY NAME	ENTITY-DOMAIN	ENTITY RANGE
— — — SHR27 used (was used by)	SHE3 SSH Project	SHE13 Work Plan
— — SHR4 had input (was input of)	SHE2 Knowledge Workflow Activity	E70 Thing
— — — SHR12 used dataset for interpretation (dataset was used by interpretation)	SHE7 Data Interpretation	SHE1 Dataset
— — — SHR31 had input for preparation (was input by preparation)	SHE6 Data Preparation and Connection	SHE1 Dataset
— — SHR17 used dataset for publication (was used by publication)	SHE8 Publication	SHE1 Dataset
— — SHR18 published	SHE8 Publication	E73 Information Object
— — SHR5 had output (was output of)	SHE2 Knowledge Workflow Activity	E70 Thing
— — — SHR10 produced dataset (dataset was produced by)	SHE5 Data Collection	SHE1 Dataset
— — — SHR11 prepared dataset (dataset was prepared by)	SHE6 Data Preparation and Connection	SHE1 Dataset
— — — SHR30 created outcome (outcome was created by)	SHE7 Data Interpretation	E73 Information Object
— — SHR6 project used (was used by project)	SHE3 SSH Project	SHE1 Dataset
— — SHR22 used tool (was tool used for)	SHE2 Knowledge Workflow Activity	SHE10 Tool
— — SHR23 used method (was method used for)	SHE2 Knowledge Workflow Activity	SHE11 Method
P94 has created (was created by)	E65 Creation	E28 Conceptual Object
— *SHR10 produced dataset (dataset was produced by)*	*SHE5 Data Collection*	*SHE1 Dataset*
P130i features are also found on	E70 Thing	E70 Thing

PROPERTY NAME	ENTITY-DOMAIN	ENTITY RANGE
— SHR20 has derivative (is derivative of)	SHE1 Dataset	SHE1 Dataset
— — SHR23 has updated version (is updated version of)	SHE1 Dataset	SHE1 Dataset
P165 incorporates (is incorporated in)	E73 Information Object	E90 Symbolic Object
— SHR21 incorporates (is incorporated in)	SHE1 Dataset	SHE1 Dataset
SHR8 used service (was service used by)	SHE2 Knowledge Workflow Activity	SHE4 Service
P134 continued (was continued by)	E7 Activity	E7 Activity
— SHR13 follows (is followed by)	SHE6 Data Preparation and Connection	SHE5 Data Collection
— SHR14 follows (is followed by)	SHE5 Data Collection	SHE7 Data Interpretation
— SHR15 follows (is followed by)	SHE6 Data Preparation and Connection	SHE7 Data Interpretation
— SHR16 follows (is followed by)	SHE7 Data Interpretation	SHE6 Data Preparation and Connection
SHR19 stored (was stored by)	SHE9 Data Storage	SHE1 Dataset

4. SSHOCro Class and Property Declaration

SSHOCro Class Declaration

The classes of the SSHOCro are comprehensively declared in this section using the following format:

- Class names are presented as headings in bold face;
- The line “Subclass of:” declares the superclass of the class from which it inherits properties;
- The line “Superclass of:” is a cross-reference to the subclasses of this class;
- The line “Scope note:” contains the textual definition of the concept the class represents;
- The line “Examples:” contains a bulleted list of examples of instances of this class. It provides illustrative examples showing how the class should be used.
- The line “Properties:” declares the list of the properties linking from a given class to other classes;
- Each property is represented by its forward name and the range class that it links to, separated by colons;

- Inherited properties are not represented;
- Properties of properties are provided indented and in parentheses beneath their respective domain property.
- Mapping to: contains the mapping to the class of the corresponding schema/model

SSHOCro Property Declaration

The properties of the CIDOC CRM are comprehensively declared in this section using the following format:

- Property names are presented as headings in bold face;
- The line "Domain:" declares the class for which the property is defined;
- The line "Range:" declares the class to which the property points, or that provides the values for the property;
- The line "Subproperty of": is a cross-reference to any superproperties the property may have;
- The line "Superproperty of:" is a cross-reference to any subproperties the property may have;
- The line "Scope note:" contains the textual definition of the concept the property represents;
- The line "Examples:" contains a bulleted list of examples of instances of this property. It provides illustrative examples showing how the property should be used.
- Mapping to: contains the mapping to the property of the corresponding schema/model

5. SSHOCro Classes

SHE1 Dataset

Subclass of: E73 Information Object

Superclass of:

Scope note: This class comprises identifiable, immaterial items such as datasets, that are documented as single information units and they may be composed of many other identifiable datasets. Datasets are a set or collection of data, records or information that is kept as a persistent unit of information in the knowledge generation process. Datasets are used as evidence for some phenomena.

This concept refers to both raw, primary data and secondary data.

Examples:

- The dataset of 50.000 definitions in Wikipedia
- ORCID Public Data File 2019
  [A snapshot of all public data in the ORCID Registry associated with an ORCID record that was created or claimed by an individual as of October 1st, 2019]
- Balafoutas & Sutter (2012) – Data.dta (Holzmeister et al., 2021a)
  [The dataset created by F. Holzmeister et.al. on September 29th 2016, to replicate the experiment of Balafoutas & Sutter (2012)]
- "daF3062e.csv" (Taloustutkimus, Suomen Kuvalehti Survey, 2015)
  [The English version of the dataset created and analysed in the context of the survey Finnish Attitudes to Immigration, commissioned by the newsmagazine Suomen Kuvalehti, in 2015]
- The spectrum "RAMAN[28]_spot[1]_R1.txt" (a dataset that was produced after scan, in a raman procedure)
- A rectified image of Holy Trinity, Anslow, Staffordshire, including specification annotations added for the conservation
- A set of annotations with relevant commentary images, result of the analysis-evaluation of the condition of a painting (which follows a macroscopic examination).

Properties: SHR20 has derivative (is derivative of): SHE1 Dataset
SHR21 incorporates (is incorporated in): SHE1 Dataset
SHR26 has updated version (is updated version of): SHE1 Dataset

Mapping to:
PE18 Dataset, SO Dataset, CERIF Result Product, MP Dataset,
HS_Dataset

SHE2 Knowledge Workflow Activity

Subclass of: E7 Activity

Superclass of: SHE5 Data Collection

SHE6 Data Preparation and Connection

SHE7 Data Interpretation

Scope note: This class comprises actions executed in a sequence to capture provenance information for improving users' knowledge, understanding and trust on the quality of data and services. Such actions support the Knowledge Creation Process, are informed by some concrete input, and result in creating or updating specific datasets and other knowledge outcomes–their outputs can also be scientific research papers, data models, specific software etc.

These processes are characterised by adherence to well-defined methods and protocols, can be iterated as a whole and may involve auxiliary processes that may form part of it, namely publication, which amounts to the activity of disseminating the outputs of a knowledge creation process and data storage, which amounts to the activities of creating back-ups for the data and other outcomes of knowledge creation process.

Examples:

- Collecting the individual participants' responses (SHE5) to the self-administered questionnaire used for the survey Finnish Attitudes to Immigration, commissioned by the newsmagazine Suomen Kuvalehti, in 2015, that took place on the 12th and the 13th of August 2015. (Taloustutkimus, Suomen Kuvalehti Survey, 2015)
- Collecting the individual participants' responses (SHE5) through face-to-face interviews, for the Swedish part of the 5th ESS Round, that took place between September 27th 2010 and March 1st 2011. (Hjerm et al. 2014)
- Calculating the overall response rate (SHE6) for the LCFS 2015-2016. (Office for National Statistics, Department for Environment, Food and Rural Affairs, 2020)
  [LCFS stands for Living Costs and Food Survey, a yearly survey whereby information is collected with regards to spending patterns and the cost of living that reflects household budgets across the UK.]

- The data manipulation process of calculating the weight variable "well-being" (SHE7), included in the 3rd edition (2019) of the LCFS 2015-2016. (Office for National Statistics, Department for Environment, Food and Rural Affairs, 2020)

Properties: SHR5 had output (was output of): E70 Thing
SHR4 had input (was input of): E70 Thing
SHR22 used tool (tool was used for): SHE10 Tool
SHR23 used method (method was used for): SHE11 Method
SHR24 based on (was basis for): SHE2 Knowledge Workflow Activity
SHR28 consists of (forms part of): SHE8 Publication
SHR29 consists of (forms part of): SHE9 Data Storage
SHR8 used service (was service used by):SHE4 Service

Mapping to: MP Activity

SHE3 SSH Project

Subclass of: E7 Activity

Superclass of:

Scope note: This class comprises instances of scientific research activities undertaken over a period of time by instances of E39 Actor. The core of such activities is an iterative process consisting of observation, rationalisation and validation phases. An instance of SHE3 SSH Project consists of the discrete knowledge workflow activities that correspond to the phases of the core process, namely: SHE5 Data Collection, SHE6 Data Preparation & Connection, and SHE7 Data Interpretation respectively. The phases can be iterated in part or as a whole, keeping track of the whole data life cycle of the research data produced and used in an instance of SHE3 SSH Project.

Examples:

- Comparative Manifesto Project
- Project P22772-G11
  [The funded project of the Austrian Science Fund which financed the research activities performed by Balafoutas and Sutter]
- The SEALIT[19] Project (ERC ID: 714437)
- The specific activities undertaken in the context of the LCFS 2015-2016 (Office for National Statistics, Department for Environment, Food and Rural Affairs, 2020).

Properties: SHR1 consists of (forms part of): SHE5 Data Collection
SHR2 consists of (forms part of): SHE6 Data Preparation and Connection
SHR3 consists of (forms part of): SHE7 Data Interpretation
SHR6 project used (was used by project): SHE1 Dataset
SHR9 has project status (is project status of): SHE12 Project Status
SHR27 used (was used by): SHE13 Work Plan

Mapping to: PE35 Project, SO Project, CERIF:Project Project, HS_Project

SHE4 Service

Subclass of: E7 Activity

Superclass of:

[19] Seafaring Lives in Transition; Mediterranean Maritime Labour and Shipping and the Challenge of Globalization: http://sealitproject.eu/?fbclid=IwAR20wxz6iHSscLK4mZGmCRBy-TxK3v0b5RHUBNVACvZixNICWyXlxu4rN8Y

Scope note: This class comprises declared offers by an instance of E39 Actor of their willingness and ability to execute some instance of E7 Activity (that may result in a product) at the request of another instance of E39 Actor, who will benefit from it.

Service depends on the individual instance of the E39 Actor making the offer, the types of activity offered and/or the types of product resulting from them.

An instance of SHE4 Service has temporal validity. Services can be temporarily interrupted without resulting in their being terminated: for example in case that the actor responsible for providing a service is on vacation or where the machine on which it relies is being repaired. It needs not continually be running in order for it to be considered continuous, for example a service may only run within certain working hours by agreement.

Examples:

- The CESSDA Vocabulary Service[20]
  [Through the CESSDA VS users can discover, browse, and download controlled vocabularies in a variety of languages.]
- The Open Science Framework (OSF) platform[21]
  [Through the OSF users can discover projects, datasets and software whereby to process them, in order to enhance their research]

Properties: SHR25 provided by (provided): E39 Actor

Mapping to: PE1 Service, CERIF:Service, MP Service

SHE5 Data Collection

Subclass of: SHE2 Knowledge Workflow Activity

E65 Creation

Superclass of:

Scope note: This class comprises the evidence collection collected by researchers directly from main sources through interviews, surveys, experiments, micro CT, etc., the gathering of objects

[20] CESSDA VS: https://vocabularies.cessda.eu
[21] Open Science Framework (OSF) platform: https://osf.io

of common criteria (qualitative and quantitative) using different methods and policies such as surveys types, research methods or by direct observation.

Examples:

- Collecting the individual participants' responses in the lab experiment undertaken by Holzmeister et al. that replicated *Balafoutas & Sutter (2012)*, as documented in Balafoutas and Sutter (2012) – Replication Report.pdf (Holzmeister et al., 2021a).
- Collecting the individual participants' responses to the self-administered questionnaire used for the survey Finnish Attitudes to Immigration, commissioned by the newsmagazine Suomen Kuvalehti, in 2015, that took place on the 12th and the 13th of August 2015. (Taloustutkimus, Suomen Kuvalehti Survey, 2015)
- Collecting the individual participants' responses to the self-administered questionnaire used for the survey Finnish Attitudes to Immigration, commissioned by the tabloid newspaper Iltalehti, in 2015, that took place between the 24th and the 27th of August 2015. (Taloustutkimus, Iltalehti Survey, 2015)
- Collecting the individual participants' responses through face-to-face interviews, for the Swedish part of the 5th ESS Round, that took place between September 27th 2010 and March 1st 2011. (Hjerm et al. 2014)
- Multispectral Acquisition Area Event 1 on 17-09-2015

Properties: SHR10 produced dataset (dataset was produced by): SHE1 Dataset
SHR14 follows (is followed by): SHE7 Data Interpretation

Mapping to:

SHE6 Data Preparation and Connection

Subclass of: SHE2 Knowledge Workflow Activity

Superclass of:

Scope note: This class comprises the construction of the contextual relationships between data, the connection, in order to provide the "latest state of knowledge". The data preparation explains the observations made and the gathering process. This procedure corrects the data that has been collected or finds lost values or converts the data to machine-readable data and analyses them (quantitative or qualitative analysis).

Examples:

- Data Entry using FastCat[22] records for the SEALIT project.
  [FAST CAT is a web-based collaborative system for assistive data entry and curation in Digital Humanities and similar forms of empirical research, built by FORTH. It is currently used by around 30 historians in the context of the SeaLiT project, a European (ERC) project of Maritime History]
- Calculating the overall response rate [46%] for the LCFS 2015-2016. (Office for National Statistics, Department for Environment, Food and Rural Affairs, 2020)
- Calculating the sample mean for variable **[q2_1]**: "How desirable or undesirable do you consider immigration by people of the following nationalities to Finland? Swedes" [3.51].
  (Taloustutkimus, Suomen Kuvalehti Survey, 2015)
  [Estimating the central tendency measures for observations collected in the context of the survey Finnish Attitudes to Immigration, commissioned by the newsmagazine Suomen Kuvalehti, in 2015]
- Calculating the sample standard deviation for variable **[q2_1]**: "How desirable or undesirable do you consider immigration by people of the following nationalities to Finland? Swedes" [0.774].
  (Taloustutkimus, Suomen Kuvalehti Survey, 2015)
  [Estimating the measures of variability for observations collected in the context of the survey Finnish Attitudes to Immigration, commissioned by the newsmagazine Suomen Kuvalehti, in 2015]
- Preparation and analysis of spectra (Raman 28 spot 1R1) after scan (part of Raman workflow).

Properties: SHR11 prepared dataset (was prepared by): SHE1 Dataset
SHR31 had input for preparation (was input by preparation): SHE1 Dataset
SHR13 follows (is followed by): SHE5 Data Collection
SHR15 follows (is followed by): SHE7 Data Interpretation

Mapping to: HS_Sample_Preparation

SHE7 Data Interpretation

Subclass of: SHE2 Knowledge Workflow Activity

Superclass of:

[22] FAST CAT: http://139.91.183.60:8181/FastCat/index.html
SeaLiT Project: http://www.sealitproject.eu/

Scope note: This class comprises the interpretation process of data, meaning the deductions by reasoning chains, the numerical evaluation of data, the comparing of data for differences or similarities, the recognition of distinct characteristics and spatiotemporal relationships between them, the inferences and the theories formulation (connecting events to a story, question events, re-examine data) and generally the process of understanding data (drawing conclusions) using arguments.

Examples:

- The data manipulation process of calculating the weight variable "well-being" (SHE7), included in the 3rd edition (2019) of the LCFS 2015-2016. (Office for National Statistics, Department for Environment, Food and Rural Affairs, 2020)
- The calculation of the probability (p=0.018) of the hypothesis that a preferential treatment of women does not render them more competitive in tournaments (H_0) supports the research hypothesis (H_a) that women do become more competitive as a result of preferential treatment. (Balafoutas & Sutter, 2012)
  The calculation of p-value (p<0.001) for the second round of replications for Duncan et al. (2012) by Holzmeister et. al (2021b) supported the replicability of experiments in the Social Sciences.
- The calculation of p-value (p=0.022) for the replication of Balafoutas & Sutter (2012) by Holzmeister et. al (2021a) supported the replicability of experiments in the Social Sciences.
- The deduction (made by A. Evans) on the existence of a pottery workshop in Knossos in the early years of the 20th century AD.
- The post processing interpretation of pigment PB15 and creation of spot results (part of the Raman spectroscopy workflow)

Properties: SHR12 used dataset for interpretation (dataset was used by interpretation): SHE1 Dataset
SHR16 follows (is followed by): SHE6 Data Preparation and Connection
SHR30 created outcome (outcome was created by): E73 Information Object

Mapping to: HS_Analysis

SHE8 Publication

Subclass of: E7 Activity

Superclass of:

Scope note: This class comprises data sharing activities (including any conditions or restrictions on that), such as releasing research papers, publication reports, datasets in order to maximise the data's discoverability.

It forms an auxiliary research activity, in the sense that it can disseminate the outputs of any research workflow activity but it does not serve to define a distinct stage thereof.

Examples:

- The activity of sharing the results of "Replication of Balafoutas and Sutter (2012)" online on OSF, on August 27th, 2018, by Holzmeister et al. (2021a).
- The activity of sharing the data for "Taloustutkimus: Finnish Attitudes to Immigration: Suomen Kuvalehti Survey 2015 [dataset]. Version 2.0 (2018-07-17). Finnish Social Science Data Archive [distributor]. http://urn.fi/urn:nbn:fi:fsd:T-FSD3062"
- The activity of releasing the 3rd edition of the LCFS 2015-2016 in 2019 (SHE8). (Office for National Statistics, Department for Environment, Food and Rural Affairs, 2020)
- The activity of publishing "Elsing, A. (2017). The Effect of Trauma and Discrimination on the Psychological Well-Being Among Asylum-Seekers in Finland. Pro gradu, Åbo Akademi." [Published research based on the data collected and processed in the context of Taloustutkimus: Finnish Attitudes to Immigration: Suomen Kuvalehti Survey 2015 [dataset]. Version 2.0 (2018-07-17). Finnish Social Science Data Archive [distributor]. http://urn.fi/urn:nbn:fi:fsd:T-FSD3062]
- The activity of publishing the Codebook for the EVA Survey on Finnish Values and Attitudes 2016.
  [The Codebook used for Finnish Business and Policy Forum (EVA): EVA Survey on Finnish Values and Attitudes 2016 [dataset]. Version 2.0 (2017-05-24). Finnish Social Science Data Archive [distributor]. http://urn.fi/urn:nbn:fi:fsd:T-FSD3093]

Properties: SHR17 used dataset for publication (was used by publication): SHE1 Dataset
SHR18 published (was published by): E73 Information Object

Mapping to: CERIF Publication

SHE9 Data Storage

Subclass of: E7 Activity

Superclass of:

Scope note: This class comprises storage activities using policies and practices for backup and storage of the data, to ensure that data will be safe and not damaged or lost.

It forms an auxiliary research activity, in the sense that it can safely store and update the outputs of any research workflow activity but it does not serve to define a distinct stage thereof.

Examples:

- The storage of the data for The Living Costs and Food Survey (UK), 2015-216 through the UK Data Service (https://ukdataservice.ac.uk/)
- The storage of the data for "Finnish Business and Policy Forum (EVA): EVA Survey on Finnish Values and Attitudes 2016 [dataset]. Version 2.0 (2017-05-24). Finnish Social Science Data Archive [distributor]. http://urn.fi/urn:nbn:fi:fsd:T-FSD3093" through the Finish Social Science Data Archive (FSD).
- Storing the dataset and relevant information (SHE9) for LCFS 2015-2016, on the UK Data Service.[23] (Office for National Statistics, Department for Environment, Food and Rural Affairs, 2020)

Properties: SHR19 stored (was stored by): SHE1 Dataset

Mapping to:

SHE10 Tool

Subclass of: E71 Human-Made Thing

Superclass of:

Scope note: This class comprises things used to perform activities. These can be material things such as a tool used for a storage activity, a recording device used in an oral data collection or an immaterial tool, such as a software.

Examples:

- STATA (statistics software)[24]
- 3M Editor[25]
- SPSS (statistics software)[26]

Properties:

Mapping to: SO Tool, CERIF:Equipment, MP Tool, HS_Device, HS_Software

[23] UK Data Service: https://ukdataservice.ac.uk/
[24] STATA: https://www.stata.com/
[25] #3M Editor: https://isl.ics.forth.gr/3M/
[26] SPSS: https://www.ibm.com/analytics/spss-statistics-software

SHE11 Method

Subclass of: E29 Design or Procedure

Superclass of:

Scope note: This class comprises the specifications, protocols, procedures, or recipes for carrying out activities.

A method may prescribe specific information resource types for inputs and outputs - methods are designed to address specific goals

Examples:

- Imputation (Bhattacherjee, 2012)
  [Missing data values treatment, according to which missing data points are substituted by values calculated according to an imputation technique; f.i. if the missing value is one item in a multi-item scale, the imputed value may be the mean of the respondent's responses to the remaining items on that scale; or, if the missing value belongs to a single-item scale, the mean of the other respondents' responses to that item can be used as the imputed value]
- Listwise deletion (Bhattacherjee, 2012)
  [Missing data values treatment, according to which, the entire observation for which missing values are reported is dropped]
- The *non-probability, purposive sampling method* that was used for Pulkkinen, Marianne (University of Jyväskylä): The Role of Local Organisations in Immigrant Integration in Jyväskylä 2018 [dataset]. Version 1.0 (2020-03-12). Finnish Social Science Data Archive [distributor]. http://urn.fi/urn:nbn:fi:fsd:T-FSD3365
- The *non-probability, respondent assisted sampling method* that was used for Pulkkinen, Marianne (University of Jyväskylä): The Role of Local Organisations in Immigrant Integration in Jyväskylä 2018 [dataset]. Version 1.0 (2020-03-12). Finnish Social Science Data Archive [distributor]. http://urn.fi/urn:nbn:fi:fsd:T-FSD3365
- Raman spectroscopy

Properties:

Mapping to: SO Method, HS_Protocol

SHE12 Project Status

Subclass of: E55 Type

Superclass of:

Scope note: This class comprises a characterization of the status of a project. A terminology is used to identify the status or the progress of a project activity.

Examples:

- ongoing
- past

Properties:

Mapping to: CERIF: Project Class

SHE13 Work Plan

Subclass of: E29 Design or Procedure

Superclass of:

Scope note: This class comprises plans foreseeing specific predefined activities or kinds of activities in the context of a project. They consist of descriptions of specific constraints, patterns or types of activities that could be realised during some time span or have the potential to be intended to be realised.

They are formal documents used to execute and control professional or business human activities with time deadlines such as project stages. They include considerations for risk management, resource management and communications, milestones, budget checks, etc.

Examples:

- D1.1 SSHOC Project Management Plan Report

Properties:

Mapping to: PE28 Curation Plan

6. SSHOCro Properties

SHR1 consists of (forms part of)

Domain: SHE3 SSH Project

Range: SHE5 Data Collection

Subproperty of:

E4 Period. P9 consists of (forms part of): E4 Period

Superproperty of:

Scope note: This property allows an instance of a Project to be analysed into parts of Data Collection activities. It is a subproperty of P9 consists of (forms part of).

This analysis of parts of the project contributes to the project data management analysis and workflow.

Examples:

- SSRP – Innsbruck: Replication of Balafoutas & Sutter (2012) (SHE3) *consists of* experiment data collection from 02/11/2016 to 24/11/2016.
  [The Social Sciences Replication Project (SSRP) performed a series of replications of experiments in the Social Sciences in order to arrive at conclusions about the robustness of experimental data-driven research, and also to assess the predictors of reproducibility for research results–see Camerer et al. 2018a, b].
- The LCFS 2015-2016 (SHE33) *consists of* calculating the overall response rate [46%] to the survey (SHE6). (Office for National Statistics, Department for Environment, Food and Rural Affairs, 2020)
- Conducting the survey Finnish Attitudes to Immigration, commissioned by the newsmagazine Suomen Kuvalehti in 2015 (SHE3) *consists of* estimating the central tendency measures for each given variable (SHE6), such as the sample mean [3.51] for **[q2_1]**: How desirable or undesirable do you consider immigration by people of the following nationalities to Finland? Swedes" (Taloustutkimus, Suomen Kuvalehti Survey, 2015)
- Conducting the survey Finnish Attitudes to Immigration, commissioned by the newsmagazine Suomen Kuvalehti in 2015 (SHE3) *consists of* estimating the measures of variability for observations collected for each given variable (SHE6), such as the sample standard deviation [0.774] for variable **[q2_1]**: "How desirable or undesirable do you consider immigration by people of the following nationalities to Finland? Swedes". (Taloustutkimus, Suomen Kuvalehti Survey, 2015)

Mapping to:

- SO Activity. SO part of: SO Activity
- MP Activity. composed of: MP Activity
- HS_Project. has_in_its_framweork (in_the_framework_of): HS_Activity

SHR2 consists of (forms part of)

Domain: SHE3 SSH Project

Range: SHE6 Data Preparation and Connection

Subproperty of:

E4 Period. P9 consists of (forms part of): E4 Period

Superproperty of:

Scope note: This property allows an instance of a Project to be analysed into Data preparation and connection activities. It is a subproperty of P9 consists of (forms part of).

This analysis of parts of the project contributes to the project data management analysis and workflow.

Examples:

- SSRP – Innsbruck: Replication of Balafoutas & Sutter (2012) (SHE3) Project *consists of* data entry (on 2017-12-20) (SHE6).
  [The Social Sciences Replication Project (SSRP) performed a series of replications of experiments in the Social Sciences in order to arrive at conclusions about the robustness of experimental data-driven research, and also to assess the predictors of reproducibility for research results–see Camerer et al. 2018a, b].
- The LCFS 2015-2016 (SHE33) *consists of* calculating the overall response rate [46%] to the survey (SHE6). (Office for National Statistics, Department for Environment, Food and Rural Affairs, 2020)
- Conducting the survey Finnish Attitudes to Immigration, commissioned by the newsmagazine Suomen Kuvalehti, in 2015 (SHE3) *consists of* estimating the central tendency measures for each given variable (SHE6), such as the sample mean [3.51] for **[q2_1]**: How desirable or undesirable do you consider immigration by people of the following nationalities to Finland? Swedes". (Taloustutkimus, Suomen Kuvalehti Survey, 2015)
- Conducting the survey Finnish Attitudes to Immigration, commissioned by the newsmagazine Suomen Kuvalehti, in 2015 (SHE3) *consists of* estimating the measures of variability for observations collected for each given variable (SHE6), such as the sample standard deviation [0.774] for variable **[q2_1]**: "How desirable or undesirable do you consider immigration by people of the following nationalities to Finland? Swedes". (Taloustutkimus, Suomen Kuvalehti Survey, 2015)

Mapping to:

- SO Activity. SO part of: SO Activity
- MP Activity. composed of: MP Activity
- HS_Project. has_in_its_framweork (in_the_framework_of): HS_Activity

SHR3 consists of (forms part of)

Domain: SHE3 SSH Project

Range: SHE7 Data Interpretation

Subproperty of:

E4 Period. P9 consists of (forms part of): E4 Period

Superproperty of:

Scope note: This property allows an instance of a Project to be analysed into Data interpretation activities. It is a subproperty of P9 consists of (forms part of).

This analysis of parts of the project contributes to the project data management analysis and workflow.

Examples:

- The LCFS 2015-2016 (SHE3) *consists of* including the data manipulation process of calculating the weight variable "well-being" (SHE7) in the 3rd edition (2019) of the LCFS 2015-2016. (Office for National Statistics, Department for Environment, Food and Rural Affairs, 2020)
- Testing the research hypothesis (H_a) that a preferential treatment of women renders them more competitive in tournaments (SHE3) consists of calculating the probability (p=0.018) that supported rejecting the null hypothesis (H_0) (SHE7). (Balafoutas & Sutter, 2012)
- The replication of Balafoutas & Sutter (2012) undertaken by the Innsbruck branch of the SSRP (SHE3) *consists of* calculating the p-value (p=0.022) that supported rejecting the null hypothesis (H_0) (SHE7), according to which replicability of experiments conducted in the Social Sciences is not feasible.
  [The Social Sciences Replication Project (SSRP) performed a series of replications of experiments in the Social Sciences in order to arrive at conclusions about the robustness of experimental data-driven research, and also to assess the predictors of reproducibility for research results–see Camerer et al. 2018a,b].
- The second round of replications for Duncan et al. (2012) undertaken by the Innsbruck branch of the SSRP (SHE3) *consists of* calculating the p-value (p<0.001) that supported rejecting the null hypothesis (H_0) (SHE7), according to which replicability of experiments conducted in the Social Sciences is not feasible. (Holzmeister et al., 2021b)
  [The Social Sciences Replication Project (SSRP) performed a series of replications of

experiments in the Social Sciences in order to arrive at conclusions about the robustness of experimental data-driven research, and also to assess the predictors of reproducibility for research results–see Camerer et al. 2018a, b].

- The deduction (made by A. Evans) on the existence of a pottery workshop in Knossos in the early years of the 20th century AD.

Mapping to:

- SO Activity. SO part of: SO Activity
- MP Activity. composed of: MP Activity
- HS_Project. has_in_its_framweork (in_the_framework_of): HS_Activity

SHR4 had input (was input of)

Domain: SHE2 Knowledge Workflow Activity

Range: E70 Thing

Subproperty of:

E7 Activity. P16 used specific object (was used for): E70 Thing

Superproperty of:

SHE7 Data Interpretation. SHR12 used dataset for interpretation (dataset was used by interpretation): SHE1 Dataset

SHE6 Data Preparation and Connection. SHR31 had input for preparation (was input by preparation): SHE1 Dataset

Scope note: This property associates an instance of a knowledge workflow activity with an instance of an object which was used as input for the specific activity.

Examples:

- The creation of the record for the sample ID No.ADNA_100062_1 (SHE5), *had input* the skeletal, left temporal bone (ID No.ADNA_100062) (E70, S13), which had been previously acquired by the Ephorate of Palaeoanthropology-Speleology from Corinth, Perachora (site 37, Π920) on April 5th, 2018. [Internal aDNA-Lab report, FORTH]
- The preparation and connection of the dataset "Ariadne aDNA list-ALL, samples_info" (SHE6) *had input* the individual records (SHE1) reporting on all previously acquired samples (such as ID No.ADNA_100062_1, ADNA_100062_2, ADNA_100062_3). [Internal aDNA-Lab report, FORTH]

Mapping to:

- MP Activity. inputEntity: MP Entity
- HS_Sample_Preparation. prepared (was_prepared_by): HS_Sample
- HS_Analysis. produced_dataset (dataset_produced_by): HS_Dataset

SHR5 had output (was output of)

Domain: SHE2 Knowledge Workflow Activity

Range: E70 Thing

Subproperty of:

E7 Activity. P16 used specific object (was used for): E70 Thing

Superproperty of:

SHE5 Data Collection. SHR10 produced dataset (dataset was produced by): SHE1 Dataset

SHE6 Data Preparation and Connection. SHR11 prepared dataset (dataset was prepared by): SHE1 Dataset

SHE7 Data Interpretation. SHR30 created outcome (outcome was created by): E73 Information Object

Scope note: This property associates an instance of a knowledge workflow activity with an instance of a thing, which was the output of the specific activity.

Examples:

- Collecting the individual participants' responses in the lab experiment undertaken by Holzmeister et al. that replicated *Balafoutas & Sutter (2012)*, from 2/11/2016 until 24/11/2016 (SHE5), *had output (SHR10)*[27] Balafoutas & Sutter (2012) – Data.dta (SHE1). (Holzmeister et al., 2021a)
- Preparation and analysis of spectra (Raman 28 spot 1R1) after scan *had output* spectrum "RAMAN[28]_spot[1]_R1.txt"

[27] The notation used (SHR10) indicates that the property is instantiated by one of its subproperties, for want of other examples.

Mapping to:

- MP Activity. outputEntity: MP Entity
- HS_Sample_Preparation. prepared (was_prepared_by): HS_Sample
- HS_Analysis. produced_dataset (dataset_produced_by): HS_Dataset

SHR6 project used (was used by project)

Domain: SHE3 SSH Project

Range: SHE1 Dataset

Subproperty of:

E7 Activity. P16 used specific object (was used for): E70 Thing

Superproperty of:

Scope note: This property describes the use of datasets in a way essential to the performance or the outcome of a SHE3 SSH Project .

Examples:

- The overall activity of replicating Balafoutas & Sutter (2012) (SHE3) *project used* Balafoutas & Sutter (2012) – Data.dta. (SHE1). (Holzmeister et al. 2021a)
- The overall activity of replicating Duncan et al. (2012) (SHE3) *project used* Duncan et al. (2012) – Data.dta. (SHE1). (Holzmeister et al. 2021a)

Mapping to:

- SO Activity. SO uses: SO InformationResource
- CERIF Project Result Product relationship

SHR8 used service (was service used by)

Domain: SHE2 Knowledge Workflow Activity

Range: SHE4 Service

Subproperty of:

Superproperty of:

Scope note: This property associates an instance of a knowledge workflow activity with the instance of the service(s) used.

Examples:

Mapping to:

- MP Activity. usedTools: MP Service

SHR9 has project status (is project status of)

Domain: SHE3 SSH Project

Range: SHE12 Project Status

Subproperty of:

E1 CRM Entity. P2 has type (is type of): E55 Type

Superproperty of:

Scope note: This property identifies the status of a project activity.

Examples:

- The Social Sciences Replication Project *has project status* concluded. [The Social Sciences Replication Project (SSRP) performed a series of replications of experiments in the Social Sciences in order to arrive at conclusions about the robustness of experimental data-driven research, and also to assess the predictors of reproducibility for research results–see Camerer et al. 2018a, b]
- The Parthenos-eu Project (SHE3) *has project status* concluded.

Mapping to:

- CERIF Project Class relationship
- MP Property. concept: MP Concept

SHR10 produced dataset (dataset was produced by)

Domain: SHE5 Data Collection

Range: SHE1 Dataset

Subproperty of:

SHE2 Knowledge Workflow Activity. SHR5 had output (was output of): E70 Thing

E65 Creation. P94 has created (was created by): E28 Conceptual Object

Superproperty of:

Scope note: This property associates an instance of a dataset with an instance of a data collection activity that produced it.

Examples:

- Collecting the individual participants' responses in the lab experiment undertaken by Holzmeister et al. that replicated *Balafoutas & Sutter (2012)*, from 2/11/2016 until 24/11/2016 (SHE5), *produced dataset* Balafoutas & Sutter (2012) – Data.dta (SHE1). (Holzmeister et al., 2021a)
- Collecting the individual participants' responses to the self-administered questionnaire used for the survey Finnish Attitudes to Immigration, commissioned by the newsmagazine Suomen Kuvalehti, in 2015, that took place on the 12th and the 13th of August 2015 produced dataset "daF3062e.csv". (Taloustutkimus, Suomen Kuvalehti Survey, 2015)

Mapping to:

- SO Activity. SO produces: SO InformationResource

SHR11 prepared dataset (was prepared by)

Domain: SHE6 Data Preparation and Connection

Range: SHE1 Dataset

Subproperty of:

SHE2 Knowledge Workflow Activity. SHR5 had output (was output of): E70 Thing

Superproperty of:

Scope note: This property associates an instance of a dataset with the instance of the data preparation and connection activity that prepared it.

Examples:

- Pre-processing the data for from the lab experiment undertaken by Holzmeister et al. that replicated *Balafoutas & Sutter (2012)* (SHE6), *prepared dataset* Balafoutas & Sutter (2012) – Data.dta (SHE1).
- Preparation and analysis of spectra (Raman 28 spot 1R1) after scan *prepared dataset* spectrum "RAMAN[28]_spot[1]_R1.txt"

Mapping to:

SHR12 used dataset for interpretation (dataset was used for interpretation)

Domain: SHE7 Data Interpretation

Range: SHE1 Dataset

Subproperty of:

SHE2 Knowledge Workflow Activity. SHR4 had input (was input of): E70 Thing

Superproperty of:

Scope note: This property identifies a dataset that was used as input by the analysis or interpretation

Examples:

- Replicability hypothesis testing by F. Holzmeister, J. Huber, M. kirchler, J. Rose *interpreted dataset* Balafoutas & Sutter (2012) – Data.dta v1.0.
- The calculation of p-value (p=0.022) for the replication of Balafoutas & Sutter (2012) by Holzmeister et. al (2021a) that supported the replicability of experiments in the Social Sciences (SHE7) interpreted dataset Balafoutas & Sutter (2012) – Data.dta. (Holzmeister et al., 2021a)
- The calculation of p-value (p<0.001) for the second round of replications for Duncan et al. (2012) by Holzmeister et. al (2021b)) that supported the replicability of experiments in the Social Sciences (SHE7) interpreted dataset Duncan et al. (2012) – Data.dta. (Holzmeister et al., 2021b)

Mapping to:

SHR13 follows (is followed by)

Domain: SHE6 Data Preparation and Connection

Range: SHE5 Data Collection

Subproperty of:

E7 Activity: P134 continued (was continued by): E7 Activity

Superproperty of:

Scope note: This property describes the sequence relation between two successive activities, where the range temporally precedes the domain. This property indicates that one instance of data preparation and connection activity follows an instance of data collection activity.

It implies a particular order between the two entities: it creates a kind of workflow (which is common in empirical research studies: data analysis follows observation and collection). Continuation implies a coherence of intentions and outcomes of the involved activities.

Examples:

- Estimating the central tendency measures for observations collected in the context of the survey Finnish Attitudes to Immigration, commissioned by the newsmagazine Suomen Kuvalehti, in 2015 (SHE6), *follows* collecting the individual participants' responses to a self-administered questionnaire (SHE5). (Taloustutkimus, Suomen Kuvalehti Survey, 2015)

Mapping to:

- SO Activity. SO follows: SO Activity
- MP Activity. follows: MP Activity

SHR14 follows (is followed by)

Domain: SHE5 Data Collection

Range: SHE7 Data Interpretation

Subproperty of:

E7 Activity: P134 continued (was continued by): E7 Activity

Superproperty of:

Scope note: This property describes the sequence relation between two activities, where the range temporally precedes the domain. This property indicates that one instance of data collection activity follows an instance of data interpretation activity.

It implies a kind of a backward order between the two entities: it creates a kind of spiral workflow process for the sake of data verification, refinement and revising, which never ends and it is always repeated.

This definition corresponds to the belief statement that all scientific research is an iterative process of observation, rationalisation, and validation.

Examples:

- The second round of data collection for the replication of Duncan et al. (2012) undertaken by the Innsbruck branch of the SSRP (SHE5) *follows* the calculation of the p-value (p=0.279>0.05) that did not support rejecting the null hypothesis (H_0) (SHE7) [according to which replicability of experiments conducted in the Social Sciences is not feasible]. (Holzmeister et al., 2021b)

Mapping to:

- SO Activity. SO follows: SO Activity
- MP Activity. follows: MP Activity

SHR15 follows (is followed by)

Domain: SHE6 Data Preparation and Connection

Range: SHE7 Data Interpretation

Subproperty of:

E7 Activity: P134 continued (was continued by): E7 Activity

Superproperty of:

Scope note: This property describes the sequence relation between two activities, where the range temporally precedes the domain. This property indicates that one instance of data preparation and connection activity follows an instance of data interpretation activity.

It implies a kind of a backward order between the two entities: it creates a kind of spiral workflow process for the sake of data verification, refinement and revising, which never ends and it is always repeated.

This definition corresponds to the belief statement that all scientific research is an iterative process of observation, rationalisation, and validation.

Examples:

- Correcting the name from "Vodowice" to "Wodowice (SHE6) *follows* the interpretation that the two person instances with names Vodowice" and "Wodowice are the same entity because they have the same birth date, death date and id (SHE7) in the context of SEALIT project.

Mapping to:

- SO Activity. SO follows: SO Activity

- MP Activity. follows: MP Activity

SHR16 follows (is followed by)

Domain: SHE7 Data Interpretation

Range: SHE6 Data Preparation and Connection

Subproperty of:

E7 Activity: P134 continued (was continued by): E7 Activity

Superproperty of:

Scope note: This property describes the sequence relation between two successive activities, where the range temporally precedes the domain. This property indicates that one instance of data interpretation activity follows an instance of data preparation and connection activity.

It implies a particular order between the two entities: it creates a kind of workflow (which is common in empirical research studies: data explanation follows data analysis). Continuation implies a coherence of intentions and outcomes of the involved activities.

Examples:

- Conclusions reached in Elsing, A. (2017). The Effect of Trauma and Discrimination on the Psychological Well-Being Among Asylum-Seekers in Finland. Pro gradu, Åbo Akademi (SHE7) *follow* estimating the measures of variability for observations collected for each given variable (SHE6) in the context of the survey Finnish Attitudes to Immigration, commissioned by the newsmagazine Suomen Kuvalehti, in 2015. (Elsing, 2017; Taloustutkimus, Suomen Kuvalehti Survey, 2015)

Mapping to:

- SO Activity. SO follows: SO Activity
- MP Activity. follows: MP Activity

SHR17 used dataset for publication (was used by publication)

Domain: SHE8 Publication

Range: SHE1 Dataset

Subproperty of:

Superproperty of:

Scope note: This property describes the use of datasets in a way essential to the performance or the outcome of a SHE8 Publication.

Examples:

- "Replication of Balafoutas and Sutter (2012)" (Online publication) (SHE8) *used dataset for publication* Balafoutas & Sutter (2012) – Data.dta (SHE1). (Holzmeister et al. 2021a)
- Publication of "Elsing, A. (2017). The Effect of Trauma and Discrimination on the Psychological Well-Being Among Asylum-Seekers in Finland. Pro gradu, Åbo Akademi." used dataset daF3062e.csv (SHE1). (Elsing, 2017)
  [Published research based on the data collected and processed in the context of Taloustutkimus: Finnish Attitudes to Immigration: Suomen Kuvalehti Survey 2015 [dataset]. Version 2.0 (2018-07-17). Finnish Social Science Data Archive [distributor]. http://urn.fi/urn:nbn:fi:fsd:T-FSD3062]

Mapping to:

- SO Activity. SO uses: SO InformationResource

SHR18 published (was published by)

Domain: SHE8 Publication

Range: E73 Information Object

Subproperty of:

Superproperty of:

Scope note: This property associates an instance of an instance of E73 Information Object (such as a research paper, monograph, technical document, 3d-models, other datasets, etc.) with the instance of SHE8 Publication that published it.

Examples:

- The Complete Repository.zip for Camerer et al. (SHE1) *was published by* its release on OSF on August 27th, 2018 (SHE8).
  [The CompleteRepository.zip contains all files used for the replications of the 21 Social Sciences experiments undertaken by the members of the SSRP]

- "daF3062e.csv" (SHE1) *was published* by the Finnish Social Science Data Archive on October 30th, 2015 (SHE8)
  ["daF3062e.csv" stands for the name of the file containing the data for "Taloustutkimus: Finnish Attitudes to Immigration: Suomen Kuvalehti Survey 2015 [dataset]. Version 2.0 (2018-07-17). Finnish Social Science Data Archive [distributor]. http://urn.fi/urn:nbn:fi:fsd:T-FSD3062" (SHE8) *published* daF3062e.csv (SHE1). (Taloustutkimus, Suomen Kuvalehti Survey, 2015)
- "Duncan et al. (2012) – Replication Report.pdf" (E73) was published by its release on OSF on August 27th, 2018 (SHE8).

Mapping to:

- SO Activity. SO produces: SO InformationResource
- CERIF Publication Result Product relationship

SHR19 stored (was stored by)

Domain: SHE9 Data Storage

Range: SHE1 Dataset

Subproperty of:

Superproperty of:

Scope note: This property associates an instance of a dataset with the data storage activity that kept and stored it.

Examples:

- The 3rd edition of the data for The Living Costs and Food Survey (UK), 2015-2016 (SHE1) *was stored by* its deposition to the UK Data Service on January 25th, 2019 (SHE9). (Office for National Statistics, Department for Environment, Food and Rural Affairs, 2020)
- "daF3062e.csv" (SHE1) *was stored by* its deposition on the Finnish Social Science Data Archive on September 23rd, 2015 (SHE9)
  ["daF3062e.csv" stands for the name of the file containing the data for "Taloustutkimus: Finnish Attitudes to Immigration: Suomen Kuvalehti Survey 2015 [dataset]. Version 2.0 (2018-07-17). Finnish Social Science Data Archive [distributor]. http://urn.fi/urn:nbn:fi:fsd:T-FSD3062" (SHE8) *published* daF3062e.csv (SHE1). (Taloustutkimus, Suomen Kuvalehti Survey, 2015)

Mapping to:

SHR20 has derivative (is derivative of)

Domain: SHE1 Dataset

Range: SHE1 Dataset

Subproperty of:

E70 Thing. P130i features are also found on (shows features of): E70 Thing

Superproperty of:

SHE1 Dataset. SHR26 has updated version (is updated version of): SHE1 Dataset

Scope note: This property associates an instance of a dataset with another instance of dataset, which modifies the content of the first one. It is a directed relationship where the domain expresses the source of derivation and the target expresses the derivative item. This property creates derivations, versions between datasets. It is a shortcut property of a derivation process (implied in the model).

Examples:

Mapping to:

- MP Dataset. related: MP Dataset

SHR21 incorporates (is incorporated in)

Domain: SHE1 Dataset

Range: SHE1 Dataset

Subproperty of:

E73 Information Object. P165 incorporates (is incorporated in): E90 Symbolic Object

Superproperty of:

Scope note: This property associates an instance of a dataset with an instance of a dataset that was included/incorporated in it.

This property makes it possible to recognize the autonomous status of the incorporated item.

Examples:

- The "Complete Repository.zip" for Camerer et al. (SHE1) *incorporates* "Balafoutas & Sutter (2012) – Data.dta (SHE1).
  [The CompleteRepository.zip contains all files used for the replications of the 21 Social Sciences experiments undertaken by the members of the SSRP. "Balafoutas & Sutter (2012) – Data.dta" is the file containing the data for the replication of Balafoutas & Sutter (2012) by Holzmeister et al. (2021a)]
- The "Complete Repository.zip" for Camerer et al. (SHE1) *incorporates* "Duncan et al. (2012) – Data.dta (SHE1).
  [The CompleteRepository.zip contains all files used for the replications of the 21 Social Sciences experiments undertaken by the members of the SSRP. "Duncan et al. (2012) – Data.dta" is the file containing the data for the replication of Duncan et al. (2012) by Holzmeister et al. (2021b)]

Mapping to:

- MP Dataset. related: MP Dataset

SHR22 used tool (tool was used for)

Domain: SHE2 Knowledge Workflow Activity

Range: SHE10 Tool

Subproperty of:

E7 Activity. P16 used specific object (was used for): E70 Thing

Superproperty of:

Scope note: This property describes the use of tools for the performance or the outcome of a Knowledge Workflow Activity.

Examples:

- Individual participants' responses to the different treatment conditions the replication of Balafoutas & Sutter (2012) *used tool* z-Tree (Camerer et al.2018b)
  [The participants' responses were registered through *z-Tree*, a piece of software especially designed for conducting economic experiments (Fischbacher, 2007)]

Mapping to:

- SO Activity. SO usesTool: SO Tool
- MP Activity. usedTools: MP Tool
- HS_Analysis. used_device (devise_used_for): HS_Device

- HS_Analysis. used_software (software_used_for): HS_Software

SHR23 used method (method was used for)

Domain: SHE2 Knowledge Workflow Activity

Range: SHE11 Method

Subproperty of:

E7 Activity. P16 used specific object (was used for): E70 Thing

Superproperty of:

Scope note: This property describes the use of methods, specifications, for carrying out a knowledge workflow activity.

Examples:

- Experiment data collection *used method* convenience sampling.
- The data collection for (Pulkkinen, 2018) (SHE5) *used* (sampling) *method* non-probability, purposive sampling (SHE11)
- The data collection for (Pulkkinen, 2018) (SHE5) *used* (sampling) *method* non-probability, respondent assisted sampling (SHE11)

Mapping to:

- SO Activity. SO employs: SO Method
- HS_Analysis. used_protocol (protocol_used_for): HS_Protocol
- HS_Activity. used_method:rdfs:Resource

SHR24 based on (was basis for)

Domain: SHE2 Knowledge Workflow Activity

Range: SHE2 Knowledge Workflow Activity

Subproperty of:

E7 Activity. P15 was influenced by (influenced): E1 CRM Entity

Superproperty of:

Scope note: This property identifies one or more knowledge workflow activities that were used as evidence, as the basis for the performance of other knowledge workflow activities.

Examples:

- The data collection for the first replication round of Duncan et al. (2012) (SHE5) *was based on* the original experiment data collection (SHE5).
- The data analysis for the second replication round of Duncan et al. (2012) (SHE5) *was based on* the original experiment data collection (SHE5).
  [The data collection for the replications undertaken in the context of SSRP followed a two-stage procedure. According to the replications protocol observed, sample sizes are calculated to have 90% power to detect 75% of the original effect size in the first round of replications and 90% power to detect 50% of the original effect size in the second round–where a second data collection takes place and the two samples are pooled. Given the original sample size of 15 participants (and observations) and standardised effect size (r= 0.674), the first round required a sample size of 36 participants, whereas the second round required an additional sample size of 56 participants–for details, see Holzmeister et al., 2021b]

Mapping to:

- MP Activity: related: MP Activity

SHR25 provided by (provided)

Domain: SHE4 Service

Range: E39 Actor

Subproperty of:

E7 Activity: P14 carried out by (performed): E39 Actor

Superproperty of:

Scope note: This property associates an instance of actor with the instance of the service that they are willing to carry out.

Examples:

Mapping to:

- PE1 Service. PP2 provided by (provides): E39 Actor
- CERIF Person Service relationship

SHR26 has updated version (is updated version of)

Domain: SHE1 Dataset

Range: SHE1 Dataset

Subproperty of:

SHE1 Dataset. SHR20 has derivative (is derivative of): SHE1 Dataset

Superproperty of:

Scope note: This property associates an instance of a dataset with another instance of dataset, which modifies and updates its content. It is a shortcut property of a version's creation process (implied in the model).

Examples:

- The released dataset "Department for Environment, Food and Rural Affairs, Office for National Statistics. (2019). *Living Costs and Food Survey, 2015-2016*. [data collection]. *3rd Edition.* UK Data Service. SN: 8210, DOI: 10.5255/UKDA-SN-8210-5" (SHE1) *is version of* "Office for National Statistics, Department for Environment, Food and Rural Affairs. (2017). *Living Costs and Food Survey, 2015-2016*. [data collection]. *2nd Edition.* UK Data Service. SN: 8210, DOI: 10.5255/UKDA-SN-8210-4" (SHE1)
  [A new release of the dataset followed the addition of the "well-being" variable to the 2015 quarter 1 raw person data].

Mapping to:

- MP Dataset. related: MP Dataset

SHR27 used (was used by)

Domain: SHE3 SSH Project

Range: SHE13 Work Plan

Subproperty of:

E7 Activity. P33 used specific technique (was used by) E29 Design or Procedure

Superproperty of:

Scope note: This property associates an instance of a Project Activity with an instance of a work plan it realised to a certain degree or at least was influenced by.

Examples:

- SSHOC Project *used* D1.1 SSHOC Project Management Plan Report.

Mapping to:

- PE3 Curating Service. PP31 uses curation plan (is curation plan used by): PE28 Curation Plan

SHR28 consists of (forms part of)

Domain: SHE2 Knowledge Workflow Activity

Range: SHE8 Publication

Subproperty of:

E4 Period. P9 consists of (forms part of): E4 Period

Superproperty of:

Scope note: This property links an instance of SHE2 Knowledge Workflow Activity (which can be a kind of data collection, data preparation or interpretation) to the SHE8 Publication activity, by means of which its results were shared with the public.

Examples:

Mapping to:

- SO Activity. SO part of: SO Activity
- MP Activity. composed of: MP Activity

SHR29 consists of (forms part of)

Domain: SHE2 Knowledge Workflow Activity

Range: SHE9 Data Storage

Subproperty of:

E4 Period. P9 consists of (forms part of): E4 Period

Superproperty of:

Scope note: This property links an instance of SHE2 Knowledge Workflow Activity (which can be a kind of data collection, data preparation or interpretation) to the SHE9 Data Storage activity, by means of which its input and output data were safely stored.

Examples:

Mapping to:

- SO Activity. SO part of: SO Activity
- MP Activity. composed of: MP Activity

SHR30 created outcome (outcome was created by)

Domain: SHE7 Data Interpretation

Range: E73 Information Object

Subproperty of:

SHE2 Knowledge Workflow Activity. SHR5 had output (was output of): E70 Thing

Superproperty of:

Scope note: This property associates an instance of an E73 Information Object with an instance of the SHE7 Data Interpretation activity that created it.

Examples:

Mapping to:

- SO Activity. SO produces: SO InformationResource

SHR31 had input for preparation (was input by preparation)

Domain: SHE6 Data Preparation and Connection

Range: SHE1 Dataset

Subproperty of:

SHE2 Knowledge Workflow Activity. SHR4 had input (was input of): E70 Thing

Superproperty of:

Scope note: This property identifies a dataset that was used as input for the preparation of a dataset

Examples:

Mapping to:

7. Referred CIDOC CRM Classes and Properties

Since the SSHOCro ontology refers to and reuses, wherever appropriate, large parts of CIDOC CRM, this section provides a comprehensive list of all relevant constructs, together with their definitions following version 7.1.1 maintained by CIDOC. Use in this context includes: reference as immediate superclass, superproperty or element of a path expression in a mapping statement.

Referred CRM Classes

This section contains the complete definitions of the classes of the CIDOC CRM Conceptual Reference Model version 7.1.1 referred to by SSHOCro ontology.

E1 CRM Entity

Superclass of: E2 Temporal Entity
E52 Time-Span
E53 Place
E54 Dimension
E77 Persistent Item
E92 Spacetime Volume

Scope note:

This class comprises all things in the universe of discourse of the CIDOC Conceptual Reference Model.

It is an abstract concept providing for three general properties:

1. Identification by name or appellation, and in particular by a preferred identifier
2. Classification by type, allowing further refinement of the specific subclass an instance belongs to
3. Attachment of free text for the expression of anything not captured by formal properties

All other classes within the CIDOC CRM are directly or indirectly specialisations of E1 CRM Entity.

Examples:

- the earthquake in Lisbon 1755 (E5) (Chester, 2001)

In First Order Logic:

E1(x)

Properties:

P1 is identified by (identifies): E41 Appellation
P2 has type (is type of): E55 Type
P3 has note: E62 String
(P3.1 has type: E55 Type)
P48 has preferred identifier (is preferred identifier of): E42 Identifier
P137 exemplifies (is exemplified by): E55 Type
(P137.1 in the taxonomic role: E55 Type)

E4 Period

Subclass of:

E2 Temporal Entity
E92 Spacetime Volume
Superclass of:
E5 Event

Scope note:

This class comprises sets of coherent phenomena or cultural manifestations occurring in time and space.

It is the social or physical coherence of these phenomena that identify an instance of E4 Period and not the associated spatiotemporal extent. This extent is only the “ground” or space in an abstract physical sense that the actual process of growth, spread and retreat has covered. Consequently, different periods can overlap and coexist in time and space, such as when a nomadic culture exists in the same area and time as a sedentary culture. This also means that overlapping land use rights, common among first nations, amounts to overlapping periods.

Often, this class is used to describe prehistoric or historic periods such as the "Neolithic Period", the "Ming Dynasty" or the "McCarthy Era", but also geopolitical units and activities of settlements are regarded as special cases of E4 Period. However, there are no assumptions about the scale of the associated phenomena. In particular all events are seen as synthetic processes consisting of coherent phenomena. Therefore, E4 Period is a superclass of E5 Event. For example, a modern clinical birth, an instance of E67 Birth, can be seen as both a single event, i.e., an instance of E5 Event, and as an extended period, i.e., an instance of E4 Period, that consists of multiple physical processes and complementary activities performed by multiple instances of E39 Actor.

As the actual extent of an instance of E4 Period in spacetime we regard the trajectories of the participating physical things during their participation in an instance of E4 Period. This includes the open spaces via which these things have interacted and the spaces by which they had the potential to interact during that period or event in the way defined by the type of the respective period or event. Examples include the air in a meeting room transferring the voices of the participants. Since these phenomena are fuzzy, we assume the spatiotemporal extent to be contiguous, except for cases of phenomena spreading out over islands or other separated areas, including geopolitical units distributed over disconnected areas such as islands or colonies.

Whether the trajectories necessary for participants to travel between these areas are regarded as part of the spatiotemporal extent or not has to be decided in each case based on a concrete analysis, taking use of the sea for other purposes than travel, such as fishing, into consideration. One may also argue that the activities to govern disconnected areas imply travelling through spaces connecting them and that these areas hence are spatially connected in a way, but it appears counterintuitive to consider for instance travel routes in international waters as extensions of geopolitical units.

Consequently, an instance of E4 Period may occupy a number of disjoint spacetime volumes, however there must not be a discontinuity in the timespan covered by these spacetime volumes. This means that an instance of E4 Period must be contiguous in time. If it has ended in all areas, it has ended as a whole. However, it may end in one area before another, such as in the Polynesian migration, and it continues as long as it is ongoing in at least one area.

We model E4 Period as a subclass of E2 Temporal Entity and of E92 Spacetime Volume. The latter is intended as a phenomenal spacetime volume as defined in CIDOC CRMgeo (Doerr & Hiebel, 2013). By virtue of this multiple inheritance, we can discuss the physical extent of an instance of E4 Period without representing each instance of it together with an instance of its associated spacetime volume. This model combines two quite different kinds of substance: an instance of E4 Period is a phenomenon while an instance of E92

Spacetime Volume is an aggregation of points in spacetime. However, the real spatiotemporal extent of an instance of E4 Period is regarded to be unique to it due to all its details and fuzziness; its identity and existence depends uniquely on the identity of the instance of E4 Period. Therefore, this multiple inheritance is unambiguous and effective and furthermore corresponds to the intuitions of natural language.

Typical use of this class in cultural heritage documentation is for documenting cultural and artistic periods. There are two different conceptualisations of 'artistic style', defined either by physical features or by historical context. For example, "Impressionism" can be viewed as a period in the European sphere of influence lasting from approximately 1870 to 1905 during which paintings with particular characteristics were produced by a group of artists that included (among others) Monet, Renoir, Pissarro, Sisley and Degas. Alternatively, it can be regarded as a style applicable to all paintings sharing the characteristics of the works produced by the Impressionist painters, regardless of historical context. The first interpretation is an instance of E4 Period, and the second defines morphological object types that fall under E55 Type.

A geopolitical unit as a specific case of an instance of E4 Period is the set of activities and phenomena related to the claim of power, the consequences of belonging to a jurisdictional area and an administrative system that establishes a geopolitical unit. Examples from the modern period are countries or administrative areas of countries such as districts whose actions and structures define activities and phenomena in the area that they intend to govern. The borders of geopolitical units are often defined in contracts or treaties although they may deviate from the actual practice. The spatiotemporal properties of Geopolitical units can be modelled through the properties inherited from E92 Spacetime Volume.

Another specific case of an instance of E4 Period is the actual extent of the set of activities and phenomena as evidenced by their physical traces that define a settlement, such as the populated period of Nineveh.

Examples:

- Jurassic (Hallam, 1975)
- Populated Period of Nineveh
- Imperial Rome under Marcus Aurelius
- European Bronze Age (Harrison, 2004)
- Italian Renaissance (Macdonald, 1992)
- Thirty Years War (Lee, 1991)
- Sturm und Drang (Berkoff, 2013)
- Cubism (Cox, 2000)

In First Order Logic:

E4(x) ⇒ E2(x)
E4(x) ⇒ E92(x)

Properties**:**

P7 took place at (witnessed): E53 Place
P8 took place on or within (witnessed): E18 Physical Thing
P9 consists of (forms part of): E4 Period

E7 Activity

Subclass of:

E5 Event

Superclass of:

E8 Acquisition
E9 Move
E10 Transfer of Custody
E11 Modification
E13 Attribute Assignment
E65 Creation
E66 Formation
E85 Joining
E86 Leaving
E87 Curation Activity

Scope note:

This class comprises actions intentionally carried out by instances of E39 Actor that result in changes of state in the cultural, social, or physical systems documented.

This notion includes complex, composite and long-lasting actions such as the building of a settlement or a war, as well as simple, short-lived actions such as the opening of a door.

Examples:

- the Battle of Stalingrad (Hoyt, 1993)
- the Yalta Conference (Harbutt, 2010)
- my birthday celebration 28-6-1995
- the writing of “Faust” by Goethe (E65) (Williams, 2020)

- the formation of the Bauhaus 1919 (E66) (Droste, 2006)
- calling the place identified by TGN '7017998' 'Quyunjig' by the people of Iraq
- Kira Weber working in glass art from 1984 to 1993
- Kira Weber working in oil and pastel painting from 1993

In First Order Logic:

E7(x) ⇒ E5(x)

Properties:

P14 carried out by (performed): E39 Actor
(P14.1 in the role of: E55 Type)
P15 was influenced by (influenced): E1 CRM Entity
P16 used specific object (was used for): E70 Thing
(P16.1 mode of use: E55 Type)
P17 was motivated by (motivated): E1 CRM Entity
P19 was intended use of (was made for): E71 Human-Made Thing
(P19.1 mode of use: E55 Type)
P20 had specific purpose (was purpose of): E5 Event
P21 had general purpose (was purpose of): E55 Type
P32 used general technique (was technique of): E55 Type
P33 used specific technique (was used by): E29 Design or Procedure
P125 used object of type (was type of object used in): E55 Type
P134 continued (was continued by): E7 Activity

E28 Conceptual Object

Subclass of:

E71 Human-Made Thing

Superclass of:

E55 Type
E89 Propositional Object
E90 Symbolic Object

Scope note:

This class comprises non-material products of our minds and other human produced data that have become objects of a discourse about their identity, circumstances of creation or historical implication. The production of such information may have been supported by the use of technical devices such as cameras or computers.

Characteristically, instances of this class are created, invented or thought by someone, and then may be documented or communicated between persons. Instances of E28 Conceptual Object have the ability to exist on more than one particular carrier at the same time, such as paper, electronic signals, marks, audio media, paintings, photos, human memories, etc.

They cannot be destroyed. They exist as long as they can be found on at least one carrier or in at least one human memory. Their existence ends when the last carrier and the last memory are lost.

Examples:

- Beethoven's "Ode an die Freude" (Ode to Joy) (E73) (Kershaw, 1999)
- the definition of "ontology" in the Oxford English Dictionary (E73) (Oxford University Press, 1989)
- the knowledge about the victory at Marathon carried by the famous runner (E89) (Lagos & Karyanos, 2020)

[Explanation note: In the following examples we illustrate the distinction between a propositional object, its names and its encoded forms. The Maxwell equations (Ball, 1962) are a good example, because they belong to the fundamental laws of physics and their mathematical content yields identical, unambiguous results regardless formulation and encoding.]

- "Maxwell equations" (E41) [preferred subject access point from LCSH, http://lccn.loc.gov/sh85082387, accessed 18th April 2021. This is only the name for the Maxwell equations as standardized by the Library of Congress and not the equations themselves.]
- "Equations, Maxwell" (E41) [variant subject access point from LCSH, http://lccn.loc.gov/sh85082387, accessed 18th April 2021. This is another name for the equation standardized by the Library of Congress and not the equations themselves.]
- Maxwell's equations (E89) [This is the propositional content of the equations proper, independent of any particular notation or mathematical formalism.] (Ball, 1962)
- The encoding of Maxwells equations as in https://upload.wikimedia.org/wikipedia/commons/c/c4/Maxwell%27sEquations.svg (E73) [accessed 18th April 2021. This is one possible symbolic encoding of the propositional content of the equations.]

In First Order Logic:

$$E28(x) \Rightarrow E71(x)$$

E29 Design or Procedure

Subclass of:

E73 Information Object

Scope note:

This class comprises documented plans for the execution of actions in order to achieve a result of a specific quality, form or contents. In particular, it comprises plans for deliberate human activities that may result in new instances of E71 Human-Made Thing or for shaping or guiding the execution of an instance of E7 Activity.

Instances of E29 Design or Procedure can be structured in parts and sequences or depend on others.

This is modelled using P69 has association with (is associated with): E29 Design or Procedure.

Designs or procedures can be seen as one of the following

1 A schema for the activities it describes
2 A schema of the products that result from their application.
3 An independent intellectual product that may have never been applied, such as Leonardo da Vinci's famous plans for flying machines.

Because designs or procedures may never be applied or only partially executed, the CIDOC CRM models a loose relationship between the plan and the respective product.

Examples:

- the ISO standardisation procedure
- the musical notation for Beethoven's "Ode to Joy" (Kershaw, 1999)
- the architectural drawings for the Kölner Dom (Cologne Cathedral) in Cologne, Germany (Wolff, 1999)
- the drawing on the folio 860 of the Codex Atlanticus from Leonardo da Vinci, 1486-1490, kept in the Biblioteca Ambrosiana in Milan

In First Order Logic:

E29(x) ⇒ E73(x)

Properties:

P68 foresees use of (use foreseen by): E57 Material
P69 has association with (is associated with): E29 Design or Procedure
(P69.1 has type: E55 Type)

E30 Right

Subclass of:

E89 Propositional Object

Scope Note:

This class comprises legal privileges concerning material and immaterial things or their derivatives.

These include reproduction and property rights.

Examples:

- copyright held by ISO on ISO/CD 21127
- ownership of the “Mona Lisa” by the museum of the Louvre, Paris, France

In First Order Logic:

E30(x) ⇒ E89(x)

E55 Type

Subclass of:

E28 Conceptual Object

Superclass of:

E56 Language

E57 Material

E58 Measurement Unit

E99 Product Type

Scope note:

This class comprises concepts denoted by terms from thesauri and controlled vocabularies used to characterize and classify instances of CIDOC CRM classes. Instances of E55 Type represent concepts in contrast to instances of E41 Appellation which are used to name instances of CIDOC CRM classes.

E55 Type is the CIDOC CRM’s interface to domain specific ontologies and thesauri. These can be represented in the CIDOC CRM as subclasses of E55 Type, forming hierarchies of terms, i.e., instances of E55 Type linked via *P127 has broader term (has narrower term)*: E55 Type. Such hierarchies may be extended with additional properties.

Examples:

- weight, length, depth [types for instances of E54]

- portrait, sketch, animation [types for instances of E36]
- French, English, German (E56)
- excellent, good, poor [types for instances of E3]
- Ford Model T, chop stick [types for instances of E22]
- cave, doline, scratch [types for instances of E26]
- poem, short story [types for instances of E33]
- wedding, earthquake, skirmish [types for instances of E5]

In First Order Logic:

E55(x) ⇒ E28(x)

Properties:

P127 has broader term (has narrower term): E55 Type

P150 defines typical parts of (define typical wholes for): E55 Type

E65 Creation

Subclass of:

E7 Activity

E63 Beginning of Existence

Superclass of:

E83 Type Creation

Scope note:

This class comprises events that result in the creation of conceptual items or immaterial products, such as legends, poems, texts, music, images, movies, laws, types etc.

Examples:

- the framing of the U.S. Constitution (Farrand, 1913)
- the drafting of U.N. resolution 1441 (United Nations Security Council, 2002)

In First Order Logic:

E65(x) ⇒ E7(x)
E65(x) ⇒ E63(x)

Properties:

P94 has created (was created by): E28 Conceptual Object

E70 Thing

Subclass of:

E77 Persistent Item

Superclass of:

E71 Human-Made Thing

E72 Legal Object

Scope note:

This general class comprises discrete, identifiable, instances of E77 Persistent Item that are documented as single units, that either consist of matter or depend on being carried by matter and are characterized by relative stability.

They may be intellectual products or physical things. They may for instance have a solid physical form, an electronic encoding, or they may be a logical concept or structure.

Examples:

- my photograph collection (E78) (fictitious)
- the bottle of milk in my refrigerator (E22) (fictitious)
- the Riss A1 plan of the Straßburger Münster (French: *Cathédrale Notre-Dame de Strasbourg*) (E29) (Liess, R., 1985)
- the thing on the top of Otto Hahn's desk (E19)
- the form of the no-smoking sign (E36)
- the cave of Dirou, Mani, Greece (E26) (Psimenos, 2005)

In First Order Logic:

E70(x) ⇒ E77(x)

Properties:

P43 has dimension (is dimension of): E54 Dimension
P101 had as general use (was use of): E55 Type
P130 shows features of (features are also found on): E70 Thing
(P130.1 kind of similarity: E55 Type)

E71 Human-Made Thing

Subclass of:

E70 Thing

Superclass of:

E24 Physical Human-Made Thing

E28 Conceptual Object

Scope note:

This class comprises discrete, identifiable human-made items that are documented as single units.

These items are either intellectual products or human-made physical things, and are characterized by relative stability. They may for instance have a solid physical form, an electronic encoding, or they may be logical concepts or structures.

Examples:

- Beethoven's 5th Symphony (E73) (Lockwood, 2015)
- Michelangelo's David (E22) (Paoletti and Bagemihl, 2015)
- Einstein's Theory of General Relativity (E89) (Hartle, 2003)
- the taxon *'Fringilla coelebs* Linnaeus,1758' (E55) (Sinkevicius and Narusevicius, 2002)

In First Order Logic:

E71(x) ⇒ E70(x)

Properties:

P102 has title (is title of): E35 Title

(P102.1 has type: E55 Type)

P103 was intended for (was intention of): E55 Type

E72 Legal Object

Subclass of:

E70 Thing

Superclass of:

E18 Physical Thing

E90 Symbolic Object

Scope note:

This class comprises those material or immaterial items to which instances of E30 Right, such as the right of ownership or use, can be applied.

This is true for all instances of E18 Physical Thing. In the case of instances of E28 Conceptual Object, however, the identity of an instance of E28 Conceptual Object or the method of its use may be too ambiguous to reliably establish instances of E30 Right, as in the case of taxa and inspirations. Ownership of corporations is currently regarded as out of scope of the CIDOC CRM.

Examples:

- the Cullinan diamond (E19) (Scarratt and Shor, 2006)
- definition of the CIDOC Conceptual Reference Model Version 5.0.4 (E73) (ISO 21127: 2014)

In First Order Logic:

E72(x) ⇒ E70(x)

Properties:

P104 is subject to (applies to): E30 Right
P105 right held by (has right on): E39 Actor

E73 Information Object

Subclass of:

E89 Propositional Object
E90 Symbolic Object

Superclass of:

E29 Design or Procedure
E31 Document
E33 Linguistic Object
E36 Visual Item

Scope note:

This class comprises identifiable immaterial items, such as poems, jokes, data sets, images, texts, multimedia objects, procedural prescriptions, computer program code, algorithm or mathematical formulae, that have an objectively recognizable structure and are documented as single units. The encoding structure known as a "named graph" also falls under this class, so that each "named graph" is an instance of E73 Information Object.

An instance of E73 Information Object does not depend on a specific physical carrier, which can include human memory, and it can exist on one or more carriers simultaneously.

Instances of E73 Information Object of a linguistic nature should be declared as instances of the E33 Linguistic Object subclass. Instances of E73 Information Object of a documentary nature should be declared as instances of the E31 Document subclass. Conceptual items such as types and classes are not instances of E73 Information Object, nor are ideas without a reproducible expression.

Examples:

- image BM000038850.JPG from the Clayton Herbarium in London (E31) (Natural History Museum, 2021)
- E. A. Poe's "The Raven" (Poe, 1869)
- the movie "The Seven Samurai" by Akira Kurosawa (Mellen, 2002)
- the text of Huray describing the Maxwell Equations (Huray, 2010)
- the Getty AAT as published as Linked Open Data, accessed 1/10/2014

In First Order Logic:

E73(x) ⇒ E89(x)

E73(x) ⇒ E90(x)

Properties:

P165 incorporates (is incorporated in): E90 Symbolic Object

E90 Symbolic Object

Subclass of:

E28 Conceptual Object
E72 Legal Object

Superclass of:

E73 Information Object
E41 Appellation

Scope note:

This class comprises identifiable symbols and any aggregation of symbols, such as characters, identifiers, traffic signs, emblems, texts, data sets, images, musical scores, multimedia objects, computer program code or mathematical formulae that have an objectively recognizable structure and that are documented as single units.

It includes sets of signs of any nature, which may serve to designate something, or to communicate some propositional content. An instance of E90 Symbolic Object may or may not have a specific meaning, for example an arbitrary character string.

In some cases, the content of an instance of E90 Symbolic Object may completely be represented by a serialized digital content model, such as a sequence of ASCII-encoded characters, an XML or HTML document, or a TIFF image. The property *P3 has note* and its subproperty *P190 has symbolic content* allow for the description of this content model. In order to disambiguate which symbolic level is the carrier of the meaning, the property *P3.1 has type* can be used to specify the encoding (e.g., "bit", "Latin character", RGB pixel).

Examples:

- 'ecognizabl'
- the "no-smoking" sign (E36)
- "BM000038850.JPG" (E41) [identifies a digital image] (Natural History Museum, 2021)
- image BM000038850.JPG from the Clayton Herbarium in London (E36) [depicts specimen of Verbesina virginica] (Natural History Museum, 2021)
- the distribution of form, tone and colour found on Leonardo da Vinci's painting named "Mona Lisa" in daylight (E36)
- the Italian text of Dante's "Divina Commedia" as found in the authoritative critical edition *La Commedia secondo l'antica vulgata a cura di Giorgio Petrocchi*, Milano: Mondadori, 1966-67 (= Le Opere di Dante Alighieri, Edizione Nazionale a cura della Società Dantesca Italiana, VII, 1-4) (Petrocchi, 1967) (E33)

In First Order Logic:

E90(x) ⇒ E28(x)

E90(x) ⇒ E72(x)

Properties:

P106 is composed of (forms part of): E90 Symbolic Object

P190 has symbolic content: E62 String

Referred CRM Properties

P2 has type (is type of)

Domain:

E1 CRM Entity

Range:

E55 Type

Superproperty of:

E1 CRM Entity. P137 exemplifies (is exemplified by): E55 Type
E13 Attribute Assignment P177 assigned property of type: E55 Type

Quantification:

many to many (0,n:0,n)

Scope note:

This property allows sub typing of CIDOC CRM entities –a form of specialisation – through the use of a terminological hierarchy, or thesaurus.

The CIDOC CRM is intended to focus on the high-level entities and relationships needed to describe data structures. Consequently, it does not specialise entities any further than is required for this immediate purpose. However, entities in the isA hierarchy of the CIDOC CRM may by specialised into any number of sub entities, which can be defined in the E55 Type hierarchy. E41 Appellation, for example, may be specialised into "e-mail address", "telephone number", "post office box", "URL" etc. none of which figures explicitly in the CIDOC CRM hierarchy. A comprehensive explanation about refining CIDOC CRM concepts by E55 Type is given in the section "About Types" in the section on "Specific Modelling Constructs" of this document.

This property is a shortcut for the path from E1 CRM Entity through *P41i was classified by,* E17 Type Assignment, *P42 assigned* to E55 Type.

Examples:

- "enquiries@cidoc-crm.org" (E41) *has type* e-mail address (E55). (fictitious)

In First Order Logic:

P2(x,y) ⇒ E1(x)
P2(x,y) ⇒ E55(y)
P2(x,y) ⇐ (∃z) [E17(z)] ∧ P41i(x,z) ∧ P42(z,y)]

P9 consists of (forms part of)

Domain:

E4 Period

Range:

E4 Period

Subproperty of:

E92 Spacetime Volume. P10i contains: E92 Spacetime Volume

Quantification:

many to many (0,n:0,n)

Scope note:

This property associates an instance of E4 Period with another instance of E4 Period that is defined by a subset of the phenomena that define the former. Therefore, the spacetime volume of the latter must fall within the spacetime volume of the former.

This property is transitive and non-symmetric.

Examples:

- Cretan Bronze Age (E4) *consists of* Middle Minoan (E4). (Hood, 1971)

In First Order Logic:

P9(x,y) ⇒ E4(x)
P9(x,y) ⇒ E4(y)
P9(x,y) ⇒ P10(y,x)
[P9(x,y) ∧ P9(y,z)] ⇒ P9(x,z)
P9(x,y) ⇒ ¬P9(y,x)

P14 carried out by (performed)

Domain:

E7 Activity

Range:

E39 Actor

Subproperty of:

E5 Event. P11 had participant (participated in): E39 Actor

Superproperty of:

E8 Acquisition. P22 transferred title to (acquired title through): E39 Actor
E8 Acquisition. P23 transferred title from (surrendered title through): E39 Actor
E10 Transfer of Custody. P28 custody surrendered by (surrendered custody through): E39 Actor
E10 Transfer of Custody. P29 custody received by (received custody through): E39 Actor

Quantification:

many to many, necessary (1,n:0,n)

Scope note:

This property describes the active participation of an instance of E39 Actor in an instance of E7 Activity.

It implies causal or legal responsibility. The *P14.1 in the role of* property of the property specifies the nature of an Actor's participation.

Examples:

- The painting of the Sistine Chapel (E7) *carried out by* Michaelangelo Buonaroti (E21) *in the role of* master craftsman (E55). (Goldscheider, 1953)

In First Order Logic:

P14(x,y) ⇒ E7(x)
P14(x,y)⇒ E39(y)
P14(x,y) ⇒ P11(x,y)
P14(x,y,z) ⇒ [P14(x,y) ∧ E55(z)]

Properties:

P14.1 in the role of: E55 Type

P15 was influenced by (influenced)

Domain:

E7 Activity

Range:

E1 CRM Entity

Superproperty of:

E7 Activity. P16 used specific object (was used for): E70 Thing
E7 Activity. P17 was motivated by (motivated): E1 CRM Entity
E7 Activity. P134 continued (was continued by): E7 Activity
E83 Type Creation. P136 was based on (supported type creation): E1 CRM Entity

Quantification:

many to many (0,n:0,n)

Scope note:

This is a high-level property, which captures the relationship between an instance of E7 Activity and anything, that is, an instance of E1 CRM Entity that may have had some bearing upon it.

The property has more specific sub properties.

Examples:

- The designing of the Sydney Harbour Bridge (E7) *was influenced by* the Tyne bridge (E22). (Dorman Long, 1932)

In First Order Logic:

P15(x,y) ⇒ E7(x)
P15(x,y) ⇒ E1(y)

P16 used specific object (was used for)

Domain:

E7 Activity

Range:

E70 Thing

Subproperty of:

E5 Event. P12 occurred in the presence of (was present at): E77 Persistent Item
E7 Activity. P15 was influenced by (influenced): E1 CRM Entity

Superproperty of:

E7 Activity. P33 used specific technique (was used by): E29 Design or Procedure
E15 Identifier Assignment. P142 used constituent (was used in): E90 Symbolic Object
E79 Part Addition. P111 added (was added by): E18 Physical Thing

Quantification:

many to many (0,n:0,n)

Scope note:

This property describes the use of material or immaterial things in a way essential to the performance or the outcome of an instance of E7 Activity.

This property typically applies to tools, instruments, moulds, raw materials and items embedded in a product. It implies that the presence of the object in question was a necessary condition for the action. For example, the activity of writing this text required the use of a computer. An immaterial thing can be used if at least one of its carriers is present. For example, the software tools on a computer.

Another example is the use of a particular name by a particular group of people over some span to identify a thing, such as a settlement. In this case, the physical carriers of this name are at least the people understanding its use.

Examples:

- The writing of the scope note of the CIDOC CRM property "P16 used specific object" contained in the CIDOC CRM version 4.1 (E7) *used specific object* Nicholas Crofts' computer (E22) *mode of use* Typing Tool; Storage Medium (E55). [the original scope note was later extended in the CIDOC CRM version 4.3]
- The people of Iraq calling the place identified by TGN '7017998' (E7) *used specific object* "Quyunjig" (E41) *mode of use* current; vernacular (E55).

In First Order Logic:

P16(x,y) ⇒ E7(x)
P16(x,y) ⇒ E70(y)
P16(x,y) ⇒ P12(x,y)
P16(x,y) ⇒ P15(x,y)
P16(x,y,z) ⇒ [P16(x,y) ∧ E55(z)]

Properties:

P16.1 mode of use: E55 Type

P33 used specific technique (was used by)

Domain:

E7 Activity

Range:

E29 Design or Procedure

Subproperty of:

E7 Activity. P16 used specific object (was used for): E70 Thing

Quantification:

many to many (0,n:0,n)

Scope note:

This property identifies a specific instance of E29 Design or Procedure in order to carry out an instance of E7 Activity or parts of it.

The property differs from *P32 used general technique (was technique of)* in that P33 refers to an instance of E29 Design or Procedure, which is a concrete information object in its own right rather than simply being a term or a method known by tradition.

Typical examples would include intervention plans for conservation or the construction plans of a building.

Examples:

- The ornamentation of silver cup 232 (E11) *used specific technique* 'Instructions for golden chase work by A N Other' (E29). (fictitious)

- The rebuilding of the German Reichstag in Berlin (E11) *used specific technique* Architectural plans by Foster and Partners (E29). (Foster, 2000)

In First Order Logic:

P33(x,y) ⇒ E7(x)
P33(x,y) ⇒ E29(y)
P33(x,y) ⇒ P16(x,y)

P94 has created (was created by)

Domain:

E65 Creation

Range:

E28 Conceptual Object

Subproperty of:

E63 Beginning of Existence. P92 brought into existence (was brought into existence by): E77 Persistent Item

Superproperty of:

E83 Type Creation. P135 created type (was created by): E55 Type

Quantification:

one to many, necessary, dependent (1,n:1,1)

Scope note:

This property links an instance of E65 Creation to the instance of E28 Conceptual Object created by it.

It represents the act of conceiving the intellectual content of the instance of E28 Conceptual Object. It does not represent the act of creating the first physical carrier of the instance of E28 Conceptual Object. As an example, this is the composition of a poem, not its commitment to paper.

Examples:

- The composition of "The Four Friends" by A. A. Milne (E65) *has created* "The Four Friends" by A. A. Milne (E33). (Milne, 2012)

In First Order Logic:

P94(x,y) ⇒ E65(x)
P94(x,y) ⇒ E28(y)
P94(x,y) ⇒ P92(x,y)

P104 is subject to (applies to)

Domain:

E72 Legal Object

Range:

E30 Right

Quantification:

many to many (0,n:0,n)

Scope note:

This property links a particular instance of E72 Legal Object to the instances of E30 Right to which it is subject.

The Right is held by an E39 Actor as described by *P75 possesses (is possessed by)*.

Examples:

- The Beatles back catalogue (E89) *is subject to* reproduction right on the Beatles back catalogue (E30). (Raga, 2016)

In First Order Logic:

P104(x,y) ⇒ E72(x)
P104(x,y) ⇒ E30(y)

P105 right held by (has right on)

Domain:

E72 Legal Object

Range:

E39 Actor

Superproperty of:

E18 Physical Thing. P52 has current owner (is current owner of): E39 Actor

Quantification:

many to many (0,n:0,n)

Scope note:

This property identifies the instance of E39 Actor who holds the instances of E30 Right to an instance of E72 Legal Object.

It is a superproperty of *P52 has current owner (is current owner of)* because ownership is a right that is held on the owned object.

This property is a shortcut of the fully developed path from E72 Legal Object, *P104 is subject to*, E30 Right, *P75i is possessed by* to E39 Actor.

Examples:

- The Beatles back catalogue (E73) *right held by* Michael Jackson (E21). (Raga, 2016)

In First Order Logic:

P105(x,y) ⇒ E72(x)
P105(x,y) ⇒ E39(y)
P105(x,y) ⇐ (∃z) [E30(z) ∧ P104(x,z) ∧ P75i(z,y)]

P106 is composed of (forms part of)

Domain:

E90 Symbolic Object

Range:

E90 Symbolic Object

Superproperty of:

E73 Information Object. P165 incorporates (is incorporated in): E90 Symbolic Object

Quantification:

many to many (0,n:0,n)

Scope note:

This property associates an instance of E90 Symbolic Object with a part of it that is by itself an instance of E90 Symbolic Object, such as fragments of texts or clippings from an image.

This property is transitive and non-reflexive.

Examples:

- This Scope note of property P106 (E33) *is composed of* 'fragments of texts' (E33).
- 'recognizable' (E90) *is composed of* 'ecognizabl' (E90).

In First Order Logic:

P106(x,y) ⇒ E90(x)
P106(x,y) ⇒ E90(y)
[P106(x,y) ∧ P106(y,z)] ⇒ P106(x,z)
¬P106(x,x)

P130 shows features of (features are also found on)

Domain:

E70 Thing

Range:

E70 Thing

Superproperty of:

E33 Linguistic Object. P73i is translation of: E33 Linguistic Object
E18 Physical Thing. P128 carries (is carried by): E90 Symbolic Object

Quantification:

many to many (0,n:0,n)

Scope note:

This property generalises the notions of "copy of" and "similar to" into a directed relationship, where the domain expresses the derivative or influenced item and the range the source or influencing item, if such a direction can be established. The property can also be used to express similarity in cases that can be stated between two objects only, without historical knowledge about its reasons. The property expresses a symmetric

relationship in case no direction of influence can be established either from evidence on the item itself or from historical knowledge. This holds in particular for siblings of a derivation process from a common source or non-causal cultural parallels, such as some weaving patterns.

The *P130.1 kind of similarity* property of the *P130 shows features of (features are also found on)* property enables the relationship between the domain and the range to be further clarified, in the sense from domain to range, if applicable. For example, it may be expressed if both items are product "of the same mould", or if two texts "contain identical paragraphs".

If the reason for similarity is a sort of derivation process, i.e., that the creator has used or had in mind the form of a particular thing during the creation or production, this process should be explicitly modelled. In these cases, *P130 shows features of* can be regarded as a shortcut of such a process. However, the current model does not contain any path specific enough to infer this property. Specializations of the CIDOC CRM may however be more explicit, for instance describing the use of moulds etc.

This property is not transitive.

Examples:

- Mary Lamb's Cymbeline from Charles and Mary Lamb's Tales from Shakespeare (E89) *shows features of* William Shakespeare's Cymbeline (E89). (Carrington, 1954)
- The audio recording of Dante Alighieri's La divina commedia read by Enrico de Negri (E73) *shows features of* the text of Dante Alighieri's La divina commedia (E89). (Alighieri, 1956)

In First Order Logic:

P130(x,y) ⇒ E70(x)
P130(x,y) ⇒ E70(y)
P130(x,y,z) ⇒ [P130(x,y) ∧ E55(z)]

Properties:

P130.1 kind of similarity: E55 Type

P165 incorporates (is incorporated in)

Domain:

E73 Information Object

Range:

E90 Symbolic Object

Subproperty of:

E90 Symbolic Object. P106 is composed of (forms part of): E90 Symbolic Object

Quantification:

many to many (0,n:0,n)

Scope note:

This property associates an instance of E73 Information Object with an instance of E90 Symbolic Object (or any of its subclasses) that was included in it.

This property makes it possible to recognise the autonomous status of the incorporated signs, which were created in a distinct context, and can be incorporated in many instances of E73 Information Object, and to highlight the difference between structural and accidental whole-part relationships between conceptual entities.

It accounts for many cultural facts that are quite frequent and significant: the inclusion of a poem in an anthology, the re-use of an operatic aria in a new opera, the use of a reproduction of a painting for a book cover or a CD booklet, the integration of textual quotations, the presence of lyrics in a song that sets those lyrics to music, the presence of the text of a play in a movie based on that play, etc.

In particular, this property allows for modelling relationships of different levels of symbolic specificity, such as the natural language words making up a particular text, the characters making up the words and punctuation, the choice of fonts and page layout for the characters.

When restricted to information objects, that is, seen as a property with E73 Information Object as domain and range the property is transitive.

A digital photograph of a manuscript page incorporates the text of a manuscript page, if the respective text is defined as a sequence of symbols of a particular type, such as Latin characters, and the resolution and quality of the digital image is sufficient to resolve these symbols so they are readable on the digital image.

Examples:

- The content of Charles-Moïse Briquet's 'Les Filigranes: dictionnaire historique des marques du papier' (E32) *incorporates* the visual aspect of the watermark used around 1358-61 by some Spanish papermaker(s) and identified as 'Briquet 4019' (E37). (Briquet, 1985)

- The visual content of Jacopo Amigoni's painting known as 'The Singer Farinelli and friends' (E36) *incorporates* the musical notation of Farinelli's musical work entitled 'La Partenza' (E73). (National Gallery of Victoria)
- The visual content of Nicolas Poussin's painting entitled 'Les Bergers d'Arcadie' (E36) *incorporates* the Latin phrase 'Et in Arcadia ego' (E33). (Wikipedia, 2020)

In First Order Logic:

P165(x,y) ⇒ E73(x)
P165(x,y) ⇒ E90(y)
P165(x,y) ⇒ P106(x,y)

8. References

Balafoutas, L. & Sutter, M. (2012). Affirmative action policies promote women and do not harm efficiency in the laboratory. *Science* **335,** 579-582. https://www.science.org/doi/10.1126/science.1211180

Barbot L., Moranville Y., Fischer F., Petitfils C., Ďurčo M., Illmayer K., Parkoła T., Wieder P., & Karampatakis S. (2019). SSHOC D7.1 System Specification – SSH Open Marketplace (1.0). Zenodo. https://doi.org/10.5281/zenodo.3547649

Bekiari, C., Bruseker, G., Doerr, M., Ore, C.E., Stead, S., Velios, A. (2021). Definition of the CIDOC Conceptual Reference Model (vol.A). https://CIDOC CRM.org/sites/default/files/cidoc_crm_v.7.1.1_0.pdf

Bhattacherjee, Al. (2012). Social Science Research: Principles, Methods, and Practices. Textbooks Collection. 3.

Broeder, D., Trippel, T., Degl'Innocenti, E., Giacomi, R., Sanesi, M., Kleemola, M., Moilanen, K., Ala-Lahti, H., Jordan, C., Alfredsson, I., L'Hours, H., & Ďurčo, M. (2019). SSHOC D3.1 Report on SSHOC (meta)data interoperability problems (v1.0). Zenodo. https://doi.org/10.5281/zenodo.3569868

Bruseker, G., Doerr, M. & Theodoridou, M. (2018). *PARTHENOS D5.5 Report on the Common Semantic Framework.* Available from Zenodo: https://zenodo.org/record/2575465#.XjQOamgzaUk

Camerer, C., Dreber, A., Holzmeister, F., Ho, T.-H., Huber, J., Johannesson, M., Kirchler, M., Nave, G., Nosek, B., Pfeiffer, T., Altmejd, A., Buttrick, N., Chan, T., Chen, Y., Forsell, E., Gampa, A., Heikenstein, E., Hummer, L., Imai, T., Isaksson, S., Manfredi, D., Rose, J., Wagenmakers, E., and Wu, H. (2018a). *Evaluating the replicability of social science experiments in Nature and Science between 2010 and 2015*. In **Nature Human Behaviour**, *2*(9), 637-644. [Retrieved February 25th, 2020, from https://osf.io/pfdyw/]

Camerer, C., Dreber, A., Holzmeister, F., Ho, T.-H., Huber, J., Johannesson, M., Kirchler, M., Nave, G., Nosek, B., Pfeiffer, T., Altmejd, A., Buttrick, N., Chan, T., Chen, Y., Forsell, E., Gampa, A., Heikenstein, E., Hummer, L., Imai, T., Isaksson, S., Manfredi, D., Rose, J., Wagenmakers, E., and Wu, H. (2018b). *Evaluating the replicability of social science experiments in Nature and Science between 2010 and 2015*; *Suplementary Infotmation*. **Open Science Framework**. [Retrieved February 25th, 2022, from https://osf.io/sva2k/]

Corti, L. & van den Eynden, V. (2015). *Learning to manage and share data: jump-starting the research methods curriculum, International Journal of Social Research Methodology*, 18:5, 545-559

DDI Alliance (2014). DDI-Codebook 2.5 XML Schema Documentation. Retrieved from: https://ddialliance.org/Specification/DDI-Codebook/2.5/XMLSchema/field_level_documentation.html [Retrieved April 4th, 2022]

Doerr, M., Bekiari, Ch., Kritsotaki, A., & Hiebel, G.H, Theodoridou, M. (2014). *Modelling Scientific Activities: Proposal for a global schema for integrating metadata about scientific observation*. CIDOC – International Documentation Committee of ICOM: Conference 6th-11th Sept. 2014 in Dresden. Available from http://network.icom.museum/cidoc/archive/past-conferences/2014-dresden/

Doerr, M., Hunter, J., Lagoze, C. (2003). *Towards a Core Ontology for Information Integration.* Journal of Digital information 4(1) (2003)

Doerr M., Iorizzo D. (2008), *The Dream of a Global Knowledge Network–A New Approach*. In Journal on Computing and Cultural Heritage (JOCCH), June 2008, Article No.: 5. https://doi.org/10.1145/1367080.1367085

Doerr M., Kritsotaki A., Christophides V., Kotzinos D. (2012). *Reference Ontology for Knowledge Creation Processes*. In: Moen A., Mørch A.I., Paavola S. (eds) Collaborative Knowledge Creation. Technology Enhanced Learning, vol 7. SensePublishers, Rotterdam.

Doerr, M., Kritsotaki, A., Rousakis, Y., Hiebel, G., Theodoridou M. (2018). *Definition of the CRMsci; An Extension of CIDOC-CRM to support scientific observation (Version 1.2.6).* [online]. Available from: http://www.cidoc-crm.org/crmsci/sites/default/files/CRMsci%20v.1.2.6.pdf [4.4.2022]

Doerr M., Theodoridou M. (2011). *CRMdig: a generic digital provenance model for scientific observation.* In: 3rd USENIX Workshop on the Theory and Practice of Provenance, TaPP 2011, Heraklion, Crete, Greece

Doerr M., Kritsotaki A., Boutsika K., (2011). Factual argumentation—a core model for assertions making. J. Comput. Cult. Herit. 3, 3, Article 8 (February 2011). https://doi.org/10.1145/1921614.1921615

Doerr M. (2003). *The CIDOC CR –An Ontological Approach to Semantic Interoperability of Metadata.* In AI Magazine, Volume.24, Number 3, pp.75-92

Duncan, K., Sadanand, A. & Davachi, L. Memory's penumbra: episodic memory decisions induce lingering mnemonic biases. *Science* **337,** 485–487 (2012). https://www.science.org/doi/10.1126/science.1221936

Ďurčo M., Barbot L., Illmayer K., Karampatakis S., Fischer F., Moranville Y., Tetteh Ocansey J., Probst S., Kozak M., Buddenbohm S., & Yim S.B. (2021). 7.2 Marketplace – Implementation (v1.0). Zenodo. https://doi.org/10.5281/zenodo.5749465

Elsing, A. (2017). The Effect of Trauma and Discrimination on the Psychological Well-Being Among Asylum-Seekers in Finland. Pro gradu, Åbo Akademi.

Finnish Business and Policy Forum (EVA). EVA Survey on Finnish Values and Attitudes 2016 [dataset]. Version 2.0 (2017-05-24). Finnish Social Science Data Archive [distributor]. http://urn.fi/urn:nbn:fi:fsd:T-FSD3093

Fischbacher, U. z-Tree: Zurich toolbox for ready-made economic experiments. *Exp.Econ.* **10,** 171–178 (2007). https://doi.org/10.1007/s10683-006-9159-4

Green, A., & Gutman, M. (2006). *Building partnerships among social science researchers, institution-based repositories and domain specific data archives.* OCLC Systems and Services: International Digital Library Perspectives, 23, 35–53.

Hjerm, M., Johansson Sevä I. (2014). ESS 5 – European Social Survey 2010, Sweden. Swedish National Data Service. Version 1.0. https://doi.org/10.5878/002220

Holzmeister F., Huber J., Kirchler M., Rose J. (2021a). *Replication of Balafoutas and Sutter (2012)*. Open Science Framework. Retrieved on March 10th 2022, from https://osf.io/m8qav/

Holzmeister, F., Huber, J., Kirchler, M., & Rose, J. (2021b). Replication of Duncan et al (2012). Retrieved on March 10th 2022, from https://osf.io/e5nvg

ISO 24622-1. (2015). Language resource management — Component Metadata Infrastructure (CMDI) — Part 1: The Component Metadata Model.

ISO 24622-2 (2019). Language resource management — Component metadata infrastructure (CMDI) — Part 2: Component metadata specification language.

ICOM/CIDOC CRM SIG (2019). *Definition of the CIDOC conceptual reference model* (version 6.2.7). Retrieved from http://www.CIDOC CRM.org/sites/default/files/CIDOC%20CRM_v6.2.7_Definition_esIP.pdf

Jeffery, K. G. et al. (2014). *Research information management: the CERIF approach.* IJMSO 9 :5-14. https://www.eurocris.org/Uploads/Web%20pages/CERIF-1.5/cerif.html

Kleemola M., Moilanen M., Broeder M., Ďurčo M., Illmayer K., Sanesi M., Degl'Innocenti E., L'Hours H., Mathers B., Fihn Marberg J., Tsoulouha E., Kritsotaki A., & Concordia C. (2021). D3.6 Report on SSHOC format interoperability solution services, including new software. https://doi.org/10.5281/zenodo.5561604

Kritsotaki A., Bekiari C., & Tsouloucha E. (2021). SSHOC Reference Ontology (SSHOCro) - SSHOC Service Catalogue's Factsheets Series (v1.0). Zenodo. https://doi.org/10.5281/zenodo.5642495

Lagoze C., Willliams J., Vilhuber L. (2013). *Encoding Provenance Metadata for Social Science Datasets.* In: Garoufallou E., Greenberg J. (eds) Metadata and Semantics Research. MTSR 2013. Communications in Computer and Information Science, vol 390. Springer, Cham

Lample K.H., Riede K., Doerr M. (2008). *Research between Natural and Cultural History Information: Benefits and IT-Requirements for Transdisciplinarity*. ACM Journal on Computing and Cultural Heritage

Leonelli, S. (2015). *What Counts as Scientific Data? A Relational Framework*. In Philosophy of science vol. 82, 5: 810-821.

Office for National Statistics, Department for Environment, Food and Rural Affairs. (2020). Living Costs and Food Survey, 2015-2016. [data collection]. 3rd Edition. UK Data Service. SN: 8210, DOI: 10.5255/UKDA-SN-8210-5

ORCID Public Data File 2019: https://doi.org/10.23640/07243.9988322.v2

Pertsas, V., Constantopoulos, P. (2017). *Scholarly Ontology: modelling scholarly practices*. In International Journal on Digital Libraries. 18, 173–190.

Pertsas, V., Constantopoulos, P. and Androutsopoulos, I. (2018). *Ontology Driven Extraction of Research Processes*. In Proceedings of the 17th International Semantic Web Conference, ISWC 2018, Monterey, CA., Oct. 2018.

Playford, C. J., Gayle, V., Connelly, R., & Gray, A. J. (2016). Administrative social science data: The challenge of reproducible research. In Big Data & Society. https://doi.org/10.1177/2053951716684143

Pulkkinen, M. (University of Jyväskylä): The Role of Local Organisations in Immigrant Integration in Jyväskylä 2018 [dataset]. Version 1.0 (2020-03-12). Finnish Social Science Data Archive [distributor]. http://urn.fi/urn:nbn:fi:fsd:T-FSD3365

Rasmussen, K.B., Blank, G. The data documentation initiative: a preservation standard for research. *Arch Sci* **7,** 55–71 (2007). https://doi.org/10.1007/s10502-006-9036-0

Richards, J., Felicetti, A., Meghini, C., & Theodoridou, M. (2020). D4.2 – Initial report on ontology implementation. Zenodo. https://doi.org/10.5281/zenodo.4916299

Romary, L., Riondet, C., Tadjou, L., Sotiropoulou, S., & Seillier, D. (2019). PARTHENOS D4.4 Report on Standardization – Final. Zenodo. https://doi.org/10.5281/zenodo.2607014

Sinclair, P., Addis, M., Choi, F., Doerr, M., Lewis, P., Martinez, K. (2006). *The use of CRM Core in Multimedia Annotation.* In SWAMM 2006. Proceedings of First International Workshop on Semantic Web Annotations for Multimedia, part of the 15th World Wide Web Conference, Edinburgh, Scotland, May 22-26, 2006

Schmidle W., Pescarin S., Demetrescu E., Fanini B., Gabellone F., Bucciero A., Chirivì A., Ferrari I., & Giuri F. (2020). D5.17 Implementation plan for the archaeological case study. Zenodo. https://doi.org/10.5281/zenodo.4558290

Taloustutkimus: Finnish Attitudes to Immigration: Iltalehti Survey 2015 [dataset]. Version 2.0 (2018-07-17). Finnish Social Science Data Archive [distributor]. http://urn.fi/urn:nbn:fi:fsd:T-FSD3063

Taloustutkimus: Finnish Attitudes to Immigration: Suomen Kuvalehti Survey 2015 [dataset]. Version 2.0 (2018-07-17). Finnish Social Science Data Archive [distributor]. http://urn.fi/urn:nbn:fi:fsd:T-FSD3062

Tsouloucha E., Kritsotaki A., Bekiari C., & Theodoridou M. (2021). D4.19 Mapping of two indicative selected standards to the SSHOCro (v1.0). Zenodo. https://doi.org/10.5281/zenodo.4457496

W3C (2011). W3C Provenance Incubator Group Wiki, Retrieved by https://www.w3.org/2005/Incubator/prov/wiki/W3C_Provenance_Incubator_Group_Wiki#About_Provenance

Wilkinson, M., Dumontier, M., Aalbersberg, I. *et al.* The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data* **3,** 160018 (2016). https://doi.org/10.1038/sdata.2016.18

Windhouwer, M., Broeder, D., & Van Uytvanck, D. (2012). A CMD core model for CLARIN web services. In Proceedings of the workshop on Describing Language Resources with Metadata: Towards Flexibility and Interoperability in the Documentation of Language Resources at LREC 2012 (pp. 41-48).

List of Figures

9. Appendix

The Raman instance used to test the mapping to SSHOCro is represented by RAMAN 28**.xml** (Link to the record:
https://isl.ics.forth.gr/FeXML-diagnosis/Index?type=RAMAN&action=view&lang=en&id=RAMAN28).
The record documents a Raman workflow, extracted from the system DIAGNOSIS.

To access the record the user should log on to the system (Link: https://isl.ics.forth.gr/diagnosis/) using the ID and password listed below.
id: username: guest,
password: guest).
The system was developed by FORTH in the context of project POLITEIA.

In what follows, the SSHOCro classes or paths corresponding to the DIAGNOSIS elements are marked in boldface. It is possible that a path rendering DIAGNOSIS element a, extends a path that has been already declared in the mapping process to express DIAGNOSIS element b. In that particular case, the full path for a. will be given, i.e., the one including b., but only the part exclusively referring to a. will be marked in boldface. For instance, DIAGNOSIS element <SpotID> is expressed in SSHOCro as follows:

- **<SpotID>**: SHE1 Dataset (–P1 is identified by: E42 Identifier)

Marked in **{curly brackets}** are the types assigned to the SSHOCro/CIDOC CRM classes to ensure that no information is lost during the mapping process. The notation, which uses the **"value"** and **"content"** along with quotation marks means that the thing that should instantiate a given class is the very string

of characters recorded in the field or subfield within the quotation marks. These specifications remain constant and have to be assigned ad hoc.

The following table provides a description of mappings for each DIAGNOSIS field to SSHOCro:

DIAGNOSIS Tag	**Unit of Information (embedding)**	**SSHOCro**
Identity of Examination	Raman/ IdentityOfExamination	**SHE2 Knowledge Workflow Activity–P2 has type: E55 Type {"Examination"}**
IdentificationNumber	Raman/ IdentityOfExamination/ IdentificationNumber	SHE2 Knowledge Workflow Activity–**P1 is identified by: E42 Identifier**
Method	Raman/ IdentityOfExamination/Method	SHE2 Knowledge Workflow Activity–**SHR23 used method: SHE11 Method**
StartingDate	Raman/ IdentityOfExamination/ StartingDate	SHE2 Knowledge Workflow Activity–**P4 has timespan: E52 Time-Span–P82 at some time within: E61 Time Primitive**
EndingDate	Raman/ IdentityOfExamination/ EndingDate	
RelatedObject	Raman/ IdentityOfExamination/ RelatedObject	SHE2 Knowledge Workflow Activity–**P12 occurred in the presence of: E70 Thing**
PersonInvolved	Raman/ IdentityOfExamination/ PersonInvolved	SHE2 Knowledge Workflow Activity–**P14 carried out by:E21 Person**
RelatedProject	Raman/ IdentityOfExamination/ RelatedProject	SHE2 Knowledge Workflow Activity–**P9i forms part of: SHE3 SSH Project**
Location	Raman/ IdentityOfExamination/ Location	SHE2 Knowledge Workflow Activity–**P7 took place at: E53 Place**
State	Raman/ IdentityOfExamination/State	SHE2 Knowledge Workflow Activity–**P2 has type: E55 Type {"State"}**
MethodReasoning	Raman/ MethodReasoning	SHE2 Knowledge Workflow Activity–**P21 had general purpose: E55 Type**
SpectrometerTest	Raman/Examination/ SpectrometerTest	SHE2 Knowledge Workflow Activity–**P9 consists of: E16 Measurement–P2 has type: E55 Type {"Spectrometer test"}**

Device	Raman/Examination/ SpectrometerTest/Device	SHE2 Knowledge Workflow Activity–P9 consists of: E16 Measurement (- P2 has type: E55 Type {"Spectrometer test"}– and **E16 Measurement.SHR22 used tool: SHE10 Tool**
ObjectInvolved	Raman/Examination/ SpectrometerTest/ ObjectInvolved	SHE2 Knowledge Workflow Activity–P9 consists of: E16 Measurement (- P2 has type: E55 Type {"Spectrometer test"}- **and E16 Measurement–P39 measured: E18 Physical Thing**
Note	Raman/Examination/ SpectrometerTest/Note	SHE2 Knowledge Workflow Activity–P9 consists of: E16 Measurement (- P2 has type: E55 Type {"Spectrometer test"} **and E16 Measurement :- P3 has note: E62 String**
Annotation	Raman/Examination/ Annotation	SHE2 Knowledge Workflow Activity–P9 consists of: **SHE6 Data Preparation and Connection: SHR11 prepared dataset: SHE1 Dataset–P2 has type: E55 Type {"Annotation"}**
SpotAnalysis	Raman/Examination/ SpotAnalysis	SHE2 Knowledge Workflow Activity–**P9 consists of: SHE5 Data Collection–P2 has type: E55 Type {"Spot Analysis"}**
SpotID	Raman/Examination/ SpotAnalysis/SpotID	SHE2 Knowledge Workflow Activity–P9 consists of: SHE5 Data Collection–P2 has type: E55 Type {"Spot Analysis"} and SHE5 Data Collection: **SHR10 produced dataset: SHE1 Dataset–P1 is identified by:E42 Identifier**
PreProcessing	Raman/Examination/ SpotAnalysis/ PreProcessing	SHE2 Knowledge Workflow Activity–P9 consists of: SHE5 Data Collection–P2 has type: E55 Type {"Spot Analysis"} and SHE5 Data Collection: **P9 consists of: SHE5 Data Collection–P2 has type: E55 Type {"PreProcessing"}**
SpotPhoto	Raman/Examination/ SpotAnalysis/ PreProcessing/ SpotInformation/ SpotPhoto	SHE2 Knowledge Workflow Activity–P9 consists of: SHE5 Data Collection–P2 has type: E55 Type {"Spot Analysis"}and and SHE5 Data Collection: P9 consists of: SHE5 Data Collection–P2 has type: E55 Type {"PreProcessing"} and **SHE5 Data Collection:SHR10 produced dataset: SHE1 Dataset :P2 has type: E55 Type {"SpotPhoto"} and** and **SHE1 Dataset:P1 is identified by: E41 Appellation**
RelatedSpotArea /Spot	Raman/Examination/ SpotAnalysis/ PreProcessing/ SpotInformation/ RelatedSpotArea/Spot	SHE2 Knowledge Workflow Activity–P9 consists of: SHE5 Data Collection–P2 has type: E55 Type {"Spot Analysis"}and and SHE5 Data Collection: P9 consists of: SHE5 Data Collection–P2 has type: E55 Type {"PreProcessing"} and SHE5 Data Collection:SHR10 produced dataset: SHE1 Dataset :P2 has type: E55 Type {"SpotPhoto"} and **SHE1 Dataset: P129 is about: SHE1 Dataset: P2 has type: E55 Type{"Spot"}**
RelatedSpotArea /Area	Raman/Examination/ SpotAnalysis/ PreProcessing/ SpotInformation/ RelatedSpotArea/Area	SHE2 Knowledge Workflow Activity–P9 consists of: SHE5 Data Collection–P2 has type: E55 Type {"Spot Analysis"}and and SHE5 Data Collection: P9 consists of: SHE5 Data Collection–P2 has type: E55 Type {"PreProcessing"} and SHE5 Data Collection:SHR10 produced dataset: SHE1 Dataset :P2 has type: E55 Type {"SpotPhoto"} **and SHE1 Dataset: P129 is about: E53 Place :P2 has type: E55 Type{"Area"}**

RelatedExamination	Raman/Examination/ SpotAnalysis/ PreProcessing/ SpotInformation/ RelatedSpotArea/ RelatedExamination	SHE2 Knowledge Workflow Activity–P9 consists of: SHE5 Data Collection–P2 has type: E55 Type {"Spot Analysis"}and and SHE5 Data Collection: P9 consists of: SHE5 Data Collection–P2 has type: E55 Type {"PreProcessing"} and SHE5 Data Collection:SHR10 produced dataset: SHE1 Dataset :P2 has type: E55 Type {"SpotPhoto"} and SHE1 Dataset: **P129 is about: SHE1 Dataset: SHR10i dataset was produced by :SHE5 Data Collection: P2 has type: E55 Type{"Examination"}**
SurfaceDescription	Raman/Examination/ SpotAnalysis/ PreProcessing/ SpotInformation/ SurfaceDescription	SHE2 Knowledge Workflow Activity–P9 consists of: SHE5 Data Collection–P2 has type: E55 Type {"Spot Analysis"}and and SHE5 Data Collection: P9 consists of: SHE5 Data Collection–P2 has type: E55 Type {"PreProcessing"} and SHE5 Data Collection:SHR10 produced dataset: SHE1 Dataset :P2 has type: E55 Type {"SpotPhoto"} and- **SHE1 Dataset :P3 has note: E62 String**
Position	Raman/Examination/ SpotAnalysis/ PreProcessing/ SpotInformation/ Position	SHE2 Knowledge Workflow Activity–P9 consists of: SHE5 Data Collection–P2 has type: E55 Type {"Spot Analysis"}and and SHE5 Data Collection: P9 consists of: SHE5 Data Collection–P2 has type: E55 Type {"PreProcessing"} and **SHE5 Data Collection:P7 took place at:E53 Place: P1 is identified by:E41 Appellation**
Color	Raman/Examination/ SpotAnalysis/ PreProcessing/ SpotInformation/Color	SHE2 Knowledge Workflow Activity–P9 consists of: SHE5 Data Collection–P2 has type: E55 Type {"Spot Analysis"}and and SHE5 Data Collection: P9 consists of: SHE5 Data Collection–P2 has type: E55 Type {"PreProcessing"} and **SHE5 Data Collection: P16 used specific object:E26 Physical Feature P2 has type: E55 Type {"Color"}**
Settings	Raman/Examination/ SpotAnalysis/ PreProcessing/Settings	SHE2 Knowledge Workflow Activity–P9 consists of: SHE5 Data Collection–P2 has type: E55 Type {"Spot Analysis"}and and SHE5 Data Collection: P9 consists of: SHE5 Data Collection–P2 has type: E55 Type {"PreProcessing"} and **SHE5 Data Collection: SHR23 used method: SHE11 Method :P2 has type: E55 Type {"protocol for settings"}**
ExposureTime	Raman/Examination/ SpotAnalysis/ PreProcessing/ Settings/ExposureTime	SHE2 Knowledge Workflow Activity–P9 consists of: SHE5 Data Collection–P2 has type: E55 Type {"Spot Analysis"}and and SHE5 Data Collection: P9 consists of: SHE5 Data Collection–P2 has type: E55 Type {"PreProcessing"} and SHE5 Data Collection: SHR23 used method: SHE11 Method :P2 has type: E55 Type {"protocol for settings"} **and SHE11 Method :P67 refers to: E52 Time-Span: P2 has type: E55 Type {"exposure time"}**
Filter	Raman/Examination/ SpotAnalysis/ PreProcessing/ Settings/Filter	SHE2 Knowledge Workflow Activity–P9 consists of: SHE5 Data Collection–P2 has type: E55 Type {"Spot Analysis"}and and SHE5 Data Collection: P9 consists of: SHE5 Data Collection–P2 has type: E55 Type {"PreProcessing"} and SHE5 Data Collection: SHR23 used method: SHE11 Method :P2 has type: E55 Type {"protocol for settings"} **and SHE11 Method :P67 refers to:E1 CRM Entity: P2 has type: E55 Type {"filter"}**

Lens	Raman/Examination/ SpotAnalysis/ PreProcessing/ Settings/Lens	SHE2 Knowledge Workflow Activity–P9 consists of: SHE5 Data Collection–P2 has type: E55 Type {"Spot Analysis"}and and SHE5 Data Collection: P9 consists of: SHE5 Data Collection–P2 has type: E55 Type {"PreProcessing"} and SHE5 Data Collection: SHR23 used method: SHE11 Method :P2 has type: E55 Type {"protocol for settings"} **and SHE11 Method :P67 refers to:E1 CRM Entity: P2 has type: E55 Type {"lens"}**
Accumulations	Raman/Examination/ SpotAnalysis/ PreProcessing/ Settings/Accumulations	SHE2 Knowledge Workflow Activity–P9 consists of: SHE5 Data Collection–P2 has type: E55 Type {"Spot Analysis"}and and SHE5 Data Collection: P9 consists of: SHE5 Data Collection–P2 has type: E55 Type {"PreProcessing"} and SHE5 Data Collection: SHR23 used method: SHE11 Method :P2 has type: E55 Type {"protocol for settings"} **and SHE11 Method :P67 refers to: E52 Time-Span: P2 has type: E55 Type {"accumulation time"}**
MeasurementOutcome	Raman/Examination/ SpotAnalysis/ MeasurementOutcome	SHE2 Knowledge Workflow Activity–P9 consists of: SHE5 Data Collection–P2 has type: E55 Type {"Spot Analysis"}and and SHE5 Data Collection: P9 consists of: SHE5 Data Collection–P2 has type: E55 Type {"PreProcessing"} and **SHE5 Data Collection :SHR13i is followed by: SHE6 Data Preparation and Connection P2 has type: E55 Type {"measurement–spectra documentation"}**
Spectrum	Raman/Examination/ SpotAnalysis/ MeasurementOutcome/Spectrum	SHE2 Knowledge Workflow Activity–P9 consists of: SHE5 Data Collection–P2 has type: E55 Type {"Spot Analysis"}and and SHE5 Data Collection: P9 consists of: SHE5 Data Collection–P2 has type: E55 Type {"PreProcessing"} and SHE5 Data Collection :SHR13iis followed by: SHE6 Data Preparation and Connection P2 has type: E55 Type {"measurement–spectra documentation"}and **SHE6 Data Preparation and Connection :SHR11 prepared dataset: SHE1 Dataset.P1 is identified by: E41 Appellation**
AfterScanPhoto	Raman/Examination/ SpotAnalysis/ MeasurementOutcome/AfterScanPhoto	SHE2 Knowledge Workflow Activity–P9 consists of: SHE5 Data Collection–P2 has type: E55 Type {"Spot Analysis"}and and SHE5 Data Collection: P9 consists of: SHE5 Data Collection–P2 has type: E55 Type {"PreProcessing"} and SHE5 Data Collection :SHR13iis followed by: SHE6 Data Preparation and Connection P2 has type: E55 Type {"measurement–spectra documentation"}and **SHE6 Data Preparation and Connection :P70i is documented in: E38 Image**
PeakOfSpectrum	Raman/Examination/ SpotAnalysis/ MeasurementOutcome/PeakOfSpectrum	SHE2 Knowledge Workflow Activity–P9 consists of: SHE5 Data Collection–P2 has type: E55 Type {"Spot Analysis"}and and SHE5 Data Collection: P9 consists of: SHE5 Data Collection–P2 has type: E55 Type {"PreProcessing"} and SHE5 Data Collection :SHR13iis followed by: SHE6 Data Preparation and Connection P2 has type: E55 Type {"measurement–spectra documentation"}and SHE6 Data Preparation and Connection :SHR11 prepared dataset: **SHE1 Dataset: P43 has dimension : E54 Dimension**

QualityOfSignal	Raman/Examination/ SpotAnalysis/ MeasurementOutcome/QualityOfSignal	SHE2 Knowledge Workflow Activity–P9 consists of: SHE5 Data Collection–P2 has type: E55 Type {"Spot Analysis"}and and SHE5 Data Collection: P9 consists of: SHE5 Data Collection–P2 has type: E55 Type {"PreProcessing"} and SHE5 Data Collection :SHR13iis followed by: SHE6 Data Preparation and Connection P2 has type: E55 Type {"measurement–spectra documentation"}and SHE6 Data Preparation and Connection :SHR11 prepared dataset: **SHE1 Dataset: P2 has type: E55 Type {"quality of signal"}**
PostProcessing	Raman/Examination/ SpotAnalysis/ PostProcessing	SHE2 Knowledge Workflow Activity–P9 consists of: SHE5 Data Collection–P2 has type: E55 Type {"Spot Analysis"}and and SHE5 Data Collection: P9 consists of: SHE5 Data Collection–P2 has type: E55 Type {"PreProcessing"} and SHE5 Data Collection :SHR13iis followed by: SHE6 Data Preparation and Connection P2 has type: E55 Type {"measurement–spectra documentation"}and **SHE6 Data Preparation and Connection : SHR16i is followed by: SHE7 Data Interpretation :P2 has type: E55 Type {"post processing**"}
Pigment	Raman/Examination/ SpotAnalysis/ PostProcessing/ Pigment	SHE2 Knowledge Workflow Activity–P9 consists of: SHE5 Data Collection–P2 has type: E55 Type {"Spot Analysis"}and and SHE5 Data Collection: P9 consists of: SHE5 Data Collection–P2 has type: E55 Type {"PreProcessing"} and SHE5 Data Collection :SHR13iis followed by: SHE6 Data Preparation and Connection P2 has type: E55 Type {"measurement–spectra documentation"}and SHE6 Data Preparation and Connection : SHR16i is followed by: SHE7 Data Interpretation :P2 has type: E55 Type {"post processing"} and **SHE7 Data Interpretation: SHR30 created outcome: SHE1 Dataset**
DatabaseName	Raman/Examination/ SpotAnalysis/ PostProcessing/ Reference/Database/ DatabaseName	SHE2 Knowledge Workflow Activity–P9 consists of: SHE5 Data Collection–P2 has type: E55 Type {"Spot Analysis"}and and SHE5 Data Collection: P9 consists of: SHE5 Data Collection–P2 has type: E55 Type {"PreProcessing"} and SHE5 Data Collection :SHR13iis followed by: SHE6 Data Preparation and Connection P2 has type: E55 Type {"measurement–spectra documentation"}and SHE6 Data Preparation and Connection : SHR16i is followed by: SHE7 Data Interpretation :P2 has type: E55 Type {"post processing"} **and SHE7 Data Interpretation: SHR22 used tool : SHE10 Tool- P2 has type: E55 Type {"database"} and SHE10 Tool: P1 is identified by:E41 Appellation**

DatabaseLink	Raman/Examination/ SpotAnalysis/ PostProcessing/ Reference/Database/ DatabaseLink	SHE2 Knowledge Workflow Activity––P9 consists of: SHE5 Data Collection–P2 has type: E55 Type {"Spot Analysis"}and and SHE5 Data Collection: P9 consists of: SHE5 Data Collection–P2 has type: E55 Type {"PreProcessing"} and SHE5 Data Collection :SHR13iis followed by: SHE6 Data Preparation and Connection P2 has type: E55 Type {"measurement–spectra documentation"}and SHE6 Data Preparation and Connection : SHR16i is followed by: SHE7 Data Interpretation :P2 has type: E55 Type {"post processing"} **and SHE7 Data Interpretation: SHR22 used tool : SHE10 Tool- P2 has type: E55 Type {"database"} and SHE10 Tool: P1 is identified by:E41 Appellation:P2 has type: E55 Type {"access point"}**
RelatedSource	Raman/Examination/ SpotAnalysis/ PostProcessing/ Reference/ RelatedSource	SHE2 Knowledge Workflow Activity–P9 consists of: SHE5 Data Collection–P2 has type: E55 Type {"Spot Analysis"}and and SHE5 Data Collection: P9 consists of: SHE5 Data Collection–P2 has type: E55 Type {"PreProcessing"} and SHE5 Data Collection :SHR13iis followed by: SHE6 Data Preparation and Connection P2 has type: E55 Type {"measurement–spectra documentation"}and SHE6 Data Preparation and Connection : SHR16i is followed by: SHE7 Data Interpretation :P2 has type: E55 Type {"post processing"} **and SHE7 Data Interpretation: P67i is referred to by:E73 Information Object**
Note	Raman/Examination/ SpotAnalysis/ PostProcessing/ Reference/Note	SHE2 Knowledge Workflow Activity–P9 consists of: SHE5 Data Collection–P2 has type: E55 Type {"Spot Analysis"}and and SHE5 Data Collection: P9 consists of: SHE5 Data Collection–P2 has type: E55 Type {"PreProcessing"} and SHE5 Data Collection :SHR13iis followed by: SHE6 Data Preparation and Connection P2 has type: E55 Type {"measurement–spectra documentation"}and SHE6 Data Preparation and Connection : SHR16i is followed by: SHE7 Data Interpretation :P2 has type: E55 Type {"post processing"} **and SHE7 Data Interpretation: P3 has note: E62 String**
SpotResults/TableOfNotes	Raman/Examination/ SpotAnalysis/ SpotResults/ TableofNotes	SHE2 Knowledge Workflow Activity–P9 consists of: SHE5 Data Collection–P2 has type: E55 Type {"Spot Analysis"}and and SHE5 Data Collection: P9 consists of: SHE5 Data Collection–P2 has type: E55 Type {"PreProcessing"} and SHE5 Data Collection :SHR13iis followed by: SHE6 Data Preparation and Connection P2 has type: E55 Type {"measurement–spectra documentation"}and SHE6 Data Preparation and Connection : SHR16i is followed by: SHE7 Data Interpretation :P2 has type: E55 Type {"post processing"} **and SHE7 Data Interpretation: P67i is referred to by:E73 Information Object**

SpotResults/Note	Raman/Examination/ SpotAnalysis/ SpotResults/Note	SHE2 Knowledge Workflow Activity–P9 consists of: SHE5 Data Collection–P2 has type: E55 Type {“Spot Analysis”}and and SHE5 Data Collection: P9 consists of: SHE5 Data Collection–P2 has type: E55 Type {“PreProcessing”} and SHE5 Data Collection :SHR13iis followed by: SHE6 Data Preparation and Connection P2 has type: E55 Type {“measurement–spectra documentation”}and SHE6 Data Preparation and Connection : SHR16i is followed by: SHE7 Data Interpretation :P2 has type: E55 Type {“post processing”} **and SHE7 Data Interpretation: P3 has note: E62 String**
File	Raman/Examination/ MethodResults/Map/ File	SHE2 Knowledge Workflow Activity–P9 consists of: SHE5 Data Collection–P2 has type: E55 Type {“Spot Analysis”}and and SHE5 Data Collection: P9 consists of: SHE5 Data Collection–P2 has type: E55 Type {“PreProcessing”} and SHE5 Data Collection :SHR13iis followed by: SHE6 Data Preparation and Connection P2 has type: E55 Type {“measurement–spectra documentation”}and SHE6 Data Preparation and Connection : SHR16i is followed by: SHE7 Data Interpretation :P2 has type: E55 Type {“post processing”} and **SHE7 Data Interpretation: SHR23 used method: SHE11 Method :P70i is documented in: E36 Visual Item :P2 has type: E55 Type {“map”} and E36 Visual Item :P1 is identified by:E41 Appellation**:
Note	Raman/Examination/ MethodResults/Map/ Note	SHE2 Knowledge Workflow Activity–P9 consists of: SHE5 Data Collection–P2 has type: E55 Type {“Spot Analysis”}and SHE5 Data Collection: P9 consists of: SHE5 Data Collection–P2 has type: E55 Type {“PreProcessing”} and SHE5 Data Collection :SHR13iis followed by: SHE6 Data Preparation and Connection P2 has type: E55 Type {“measurement–spectra documentation”}and SHE6 Data Preparation and Connection : SHR16i is followed by: SHE7 Data Interpretation :P2 has type: E55 Type {“post processing”} and **SHE7 Data Interpretation : SHR23 used method: SHE11 Method : P70i is documented in: E36 Visual Item: :P2 has type: E55 Type {“map”} and E36 Visual Item :P3 has note: E62 String**
Photo	Raman/Examination/ MethodResults/ ReportPhoto/Photo	SHE2 Knowledge Workflow Activity–P9 consists of: SHE5 Data Collection–P2 has type: E55 Type {“Spot Analysis”}and and SHE5 Data Collection: P9 consists of: SHE5 Data Collection–P2 has type: E55 Type {“PreProcessing”} and SHE5 Data Collection :SHR13iis followed by: SHE6 Data Preparation and Connection P2 has type: E55 Type {“measurement–spectra documentation”}and SHE6 Data Preparation and Connection : SHR16i is followed by: SHE7 Data Interpretation :P2 has type: E55 Type {“post processing”} and **SHE7 Data Interpretation : SHR23 used method: SHE11 Method : P70i is documented in: E36 Visual Item: :P2 has type: E55 Type {“photo”} and E36 Visual Item :P1 is identified by:E41 Appellation**:

Note	Raman/Examination/ MethodResults/ ReportPhoto/Note	SHE2 Knowledge Workflow Activity–P9 consists of: SHE5 Data Collection–P2 has type: E55 Type {“Spot Analysis”}and and SHE5 Data Collection: P9 consists of: SHE5 Data Collection–P2 has type: E55 Type {“PreProcessing”} and SHE5 Data Collection :SHR13iis followed by: SHE6 Data Preparation and Connection P2 has type: E55 Type {“measurement–spectra documentation”}and SHE6 Data Preparation and Connection : SHR16i is followed by: SHE7 Data Interpretation :P2 has type: E55 Type {“post processing”} and **SHE7 Data Interpretation : SHR23 used method: SHE11 Method :P70i is documented in: E36 Visual Item: :P2 has type: E55 Type {“photo”} and E36 Visual Item :P3 has note: E62 String**
TableofNotes	Raman/Examination/ MethodResults/ TableOfNotes	SHE2 Knowledge Workflow Activity–P9 consists of: SHE5 Data Collection–P2 has type: E55 Type {“Spot Analysis”}and and SHE5 Data Collection: P9 consists of: SHE5 Data Collection–P2 has type: E55 Type {“PreProcessing”} and SHE5 Data Collection :SHR13iis followed by: SHE6 Data Preparation and Connection P2 has type: E55 Type {“measurement–spectra documentation”}and SHE6 Data Preparation and Connection : SHR16i is followed by: SHE7 Data Interpretation :P2 has type: E55 Type {“post processing”} and **SHE7 Data Interpretation: SHR23 used method: SHE11 Method :P70i is documented in: E31 Document**
Note	Raman/Examination/ /MethodResults/Note	SHE2 Knowledge Workflow Activity–P9 consists of: SHE5 Data Collection–P2 has type: E55 Type {“Spot Analysis”}and and SHE5 Data Collection: P9 consists of: SHE5 Data Collection–P2 has type: E55 Type {“PreProcessing”} and SHE5 Data Collection :SHR13iis followed by: SHE6 Data Preparation and Connection P2 has type: E55 Type {“measurement–spectra documentation”}and SHE6 Data Preparation and Connection : SHR16i is followed by: SHE7 Data Interpretation :P2 has type: E55 Type {“post processing”} and **SHE7 Data Interpretation : SHR23 used method: SHE11 Method :P3 has note: E62 String**