Research and Innovation Action

Social Sciences & Humanities Open Cloud

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

Deliverable 4.18 SSHOCro beta version

Dissemination Level	PU
Due Date of Deliverable	31/12/20, (M12)
Actual Submission Date	26/03/20
Work Package	WP 4 - Innovations in data production
Task	T4.7 Modeling the SSHOC data life cycle
Type	OTHER
Approval Status	Waiting EC approval
Version	v1.0 (beta)
Number of Pages	p.1 – p.57

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 life cycle used by Social Science and Humanities researchers. The SSHOCro is provided in RDF/S in the file titled "D4.18 SSHOCro_v.1.0_beta.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
0.0	20/12/2019	Initial report of deliverable D4.18, classes and properties definitions - v 0.0	Athina Kritsotaki, Eleni Tsouloucha, Chryssoula Bekiari
0.1	15/01/2020	Report of deliverable D4.18, classes and properties definitions - v 1.0	Athina Kritsotaki, Eleni Tsouloucha, Chryssoula Bekiari
0.1	15/01/2020	RDF representation of deliverable D4.18, classes and properties definitions - v 1.0	Athina Kritsotaki
0.1	15/01/2020	update schemas and example	Eleni Tsouloucha, Athina Kritsotaki
0.12	20/01/2020	comments, updates, on the executive summary, introduction	Chryssoula Bekiari
0.13	05/02/2020	addition of mapping to MP data model	Athina Kritsotaki
0.14	25/02/2020	comment on KWA & data storage	Maurizio Sanesi (CNR)
0.14	20/02/2020	Peer review	Matej Ďurčo (OEAW), Joseph Padfield (NG-London), Nicolas Larrousse (huma-num.CNRS)
1.0	28/02/2020	Address peer review comments	Athina Kritsotaki, Eleni Tsouloucha, Chryssoula Bekiari

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

Executive Summary

This document serves as a beta (v 1.0) of the definition of the SSHOC Reference Ontology (SSHOCro). SSHOCro proposes an ontological model and RDF schema to be used as a top-level ontology for organizing 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 actually 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 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 to a set of activities performed in it.

The proposed ontology is based on the following considerations:

1. Cultural and scientific data cannot be understood without knowledge about the meaning of the data and the ways and circumstances of their creation. This knowledge comprises the provenance of 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 artifact. (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. The provenance metadata are actually data describing objects, people, places, times which are causally related by events while other relations are either deductions from events or found by observation events. (Doerr & Theodoridou, 2011)

4. Provenance metadata are event centric and must be described in a historical order in order to ensure that there are no references to non-existent (non-recorded) events or objects. Generally, metadata is used to assess meaning (view, experimental setup, instrument settings), relevance (depicted objects, their status, their conditions), quality (calibration, tolerances, errors, artifacts) and possibilities of improvement and reprocessing. (Doerr & Theodoridou, 2011)

5. 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)

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

7. The social science & humanities research is a repetitive process of (i) formulating questions (ii) finding empirical evidence (iii) interpretation (inference, causation) (iv) verification in wider context (v) triggering new questions (Doerr et al., 2011)

8. From a data processing point of view in the social science and humanities research there is a dominating "collection, connection, interpretation" pattern having 3 auxiliary activities concerning (i) the 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 (iii) Information Selection, employing finding, retrieving, inspecting, and selecting actions

SSHOCro is modelled as an extension of CIDOC CRM, the ISO standard ontology for Cultural Heritage data, from which it inherits its event-centric orientation. 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 for good practice of conceptual modeling. 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 SSHOCro is provided in RDF/S in the SSHOCro_v.1.0_beta.rdf, which is attached to this report.

Abbreviations and Acronyms

CERIF	Common European Research Information Format Model
CIDOC CRM	CIDOC Conceptual Reference Model
CRMsci	Scientific Observation Model
EOSC	European Open Science Cloud
MP	Marketplace
OSF	Open Science Framework
PARTHENOS	Pooling Activities, Resources and Tools for Heritage E-research Networking, Optimization and Synergies
RDF/S	Resource Description Framework Schema
SEALIT	Seafaring Lives in Transition
SSH	Social Sciences and Humanities
SSHOC	Social Sciences and Humanities Open Cloud
SSHOCro	SSHOC Reference Ontology
SO	Scholarly Ontology
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 organizing 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 in order to describe at a generic level the real-world lifecycle of data produced by the generic workflow of processes of collection, connection, interpretation and 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 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 frame, 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 are the ensuing data 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 in it.

SSHOCro is being 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.

In the remainder of this document, the reader can navigate through the classes and properties of the SSHOCro.

1.2 Relation to CIDOC CRM

SSHOCro uses and extends the CIDOC CRM (ISO21127), a fundamental ontology which aims at linking and integrating information relating to cultural heritage, based on a common conceptual approach. CIDOC CRM focuses on event modeling in order to describe cultural materials and scientific observations and it provides a uniform and extensible semantic framework that can gather, link and integrate any information relating to cultural heritage. The purpose of this standard is to provide a common language among experts in the field (historians, archaeologists, conservators, curators, etc.) to formulate requirements for information systems, to serve as a guide of good practice for conceptual modeling and to be a core schema for the development of

large knowledge networks connected to the Web. Apart from the use of cultural heritage, its usefulness in applications in 'electronic' science, (e-science) and biodiversity has been proven.

The CIDOC CRM is a living standard that is designed in such a way as to provide both high level information retrieval and the formulation and documentation of very specific data points and questions. The CIDOC CRM thus consists of the CRMbase standard which provides the basic classes and relations devised for the cultural heritage world. This base ontology is complemented by a series of modular extensions to the basic model. Such extensions are designed ***to support different types of specialized research questions and documentation*** such as bibliographic documentation or geoinformatics or archaeological excavations etc. The CIDOC CRM extensions are developed in partnership with the research communities in question. These extensions are formulated in a manner that is harmonized with the base ontology such that data expressed in any extension is compatible with the base system of concepts and relations. This harmonized development process leads to a high level of information integrity and integration not available in other information systems.

The proposed SSHOCro ontology is modelled as an extension of CIDOC CRM which addresses the dominating repetitive pattern "collection, connection, interpretation" found in SSH research along with its auxiliary activities concerning the persistent storage, publication & presentation as well as information selection. It also builds on existing CIDOC CRM extensions, such as CRMsci (a model for integrating metadata concerning scientific observation, measurements and processed data in descriptive and empirical sciences) and the PARTHENOS Entities (a model to support the activities and aims of Research Infrastructures). Given its open extension policy and support of analytic data generated by empirical sciences, CIDOC CRM provides a suitably general ontology to support the integration of data and metadata across the Social Sciences and the Humanities.

1.3 Naming Convention

SSHOCro classes and properties linking them to one another will be given both a name and an identifier following the conventions of the CIDOC CRM. However, in this beta version, SSHOCro classes and properties are not assigned identifiers but only labels. It will be assigned proper identifiers in subsequent finalized versions. Thus, the following naming conventions have been applied throughout the proposed SSHOCro reference ontology.

- Class labels are preceded by the letters “SHE” (stands for Social science and Humanities Entity-historically classes were sometimes referred to as “Entities”) and are named using noun phrases (nominal groups) using title case (initial capitals). For example, SHE Project Activity.
- Property labels are preceded by the letters “SHR” (stands for Social science and Humanities Relationship). For example, SHR follows.
- Property names should be read in their non-parenthetical form for the domain-to-range direction, and in 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 letters "**SHE**" or "**SHR**" of the labels of the classes and properties are followed by “**i**” respectively. For example, ***SHRi data interpretation part of*** links from SHE Data

Interpretation to SHE Project Activity, in the inverse order from the forward going property, SHR consists of data interpretation.

1.4 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 structure of E7 Activity, following the event modelling, realizes an event pattern[1], which is a strong mechanism for integrating heterogeneous complementary information. As such, instances of E7 Activity link actions to the actors that committed them, the timespan they unfolded over, the place(s) where they transpired, the manner they unfolded and the changes they brought about in the world (or that they resulted from) – see **Figure 1**, below.

[1] An advantage of its use is that all the data elements following this pattern can directly be mapped to the CIDOC CRM.

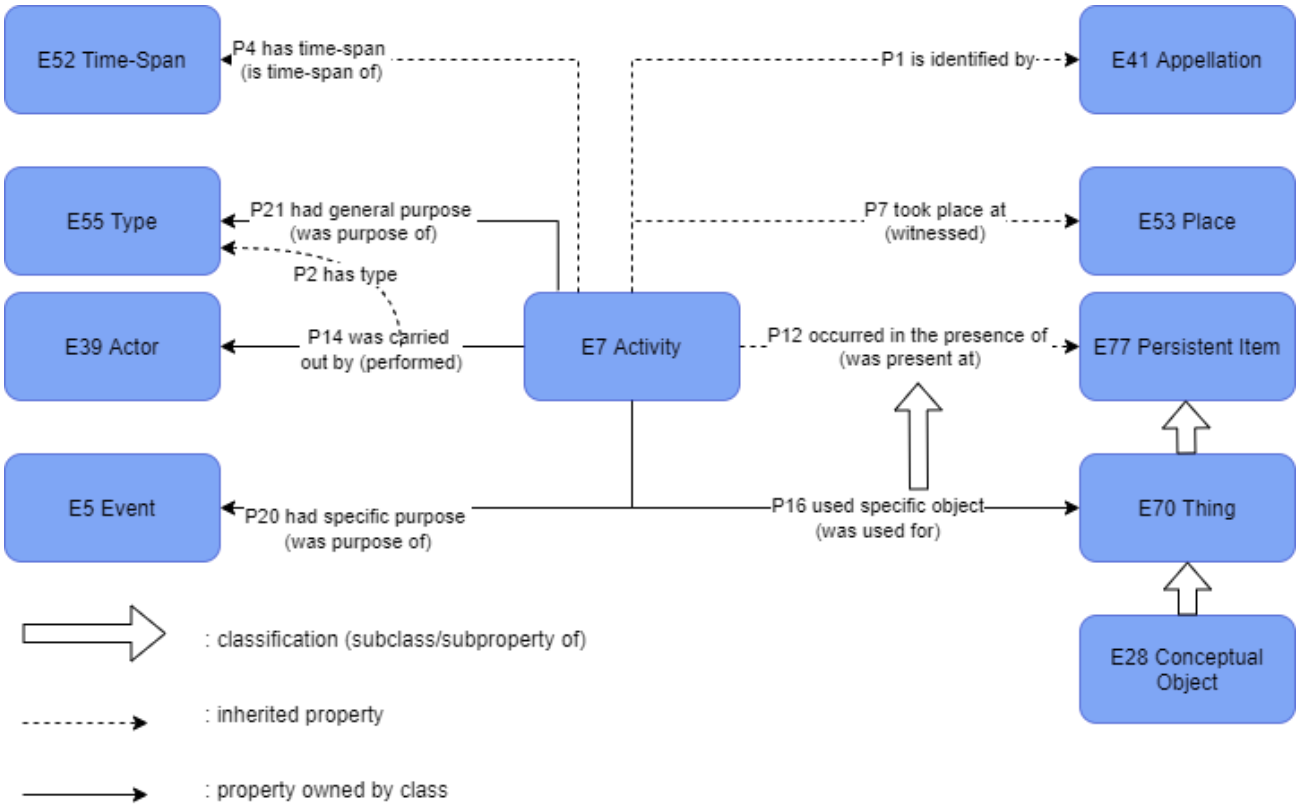

Figure 1: CIDOC CRM E7 Activity

Of the entities represented in the diagram, **Things** correspond to discrete, identifiable, persistent items. They can be material –like any sort of concrete object –or immaterial –f.i. images, texts, datasets, organizational 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 characterize 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 they are involved in, affecting the things they interact with 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. The diagram above only represents part-decomposition with respect to Activities, as this is considered more relevant for SSHOCro.

Causal relations like influence and motivation for an activity or an event can also be documented with the CIDOC CRM.

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 diagrams. The first of them represents the temporal entities of SSHOCro –all subclasses of E7 Activity; the second 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.

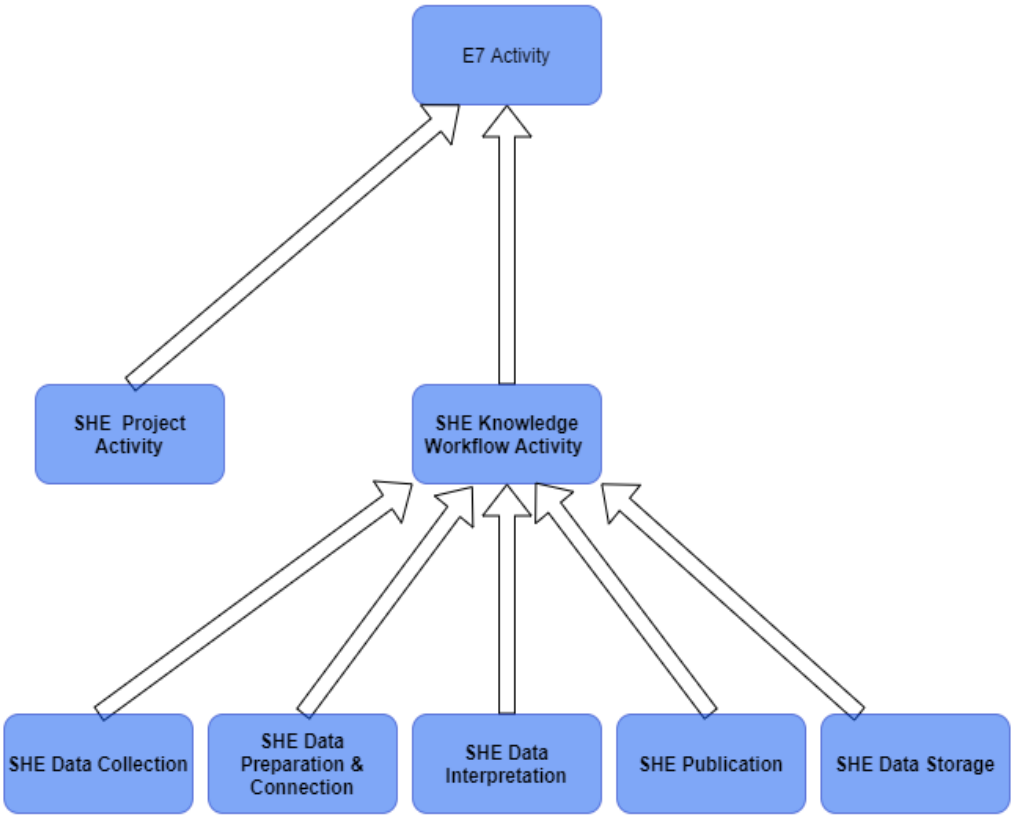

Figure 2: SSHOCro temporal entities

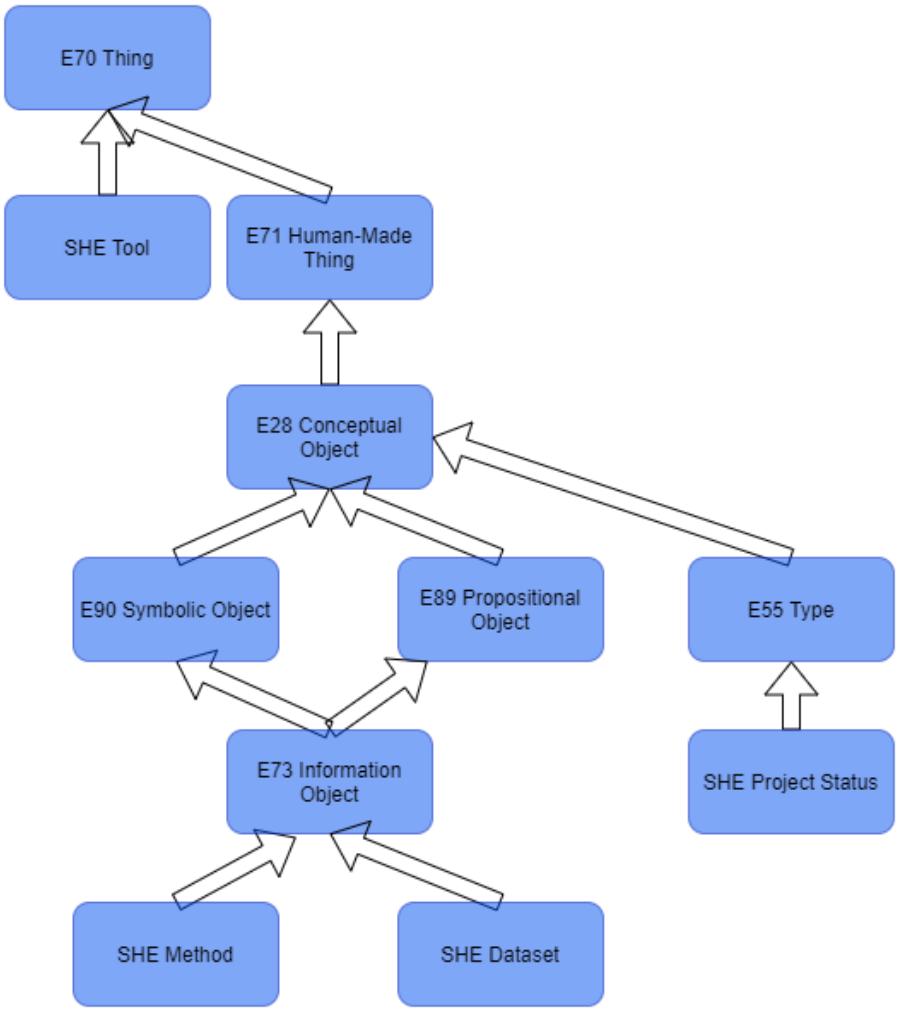

Figure 3: SSHOCro Things

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 information objects as inputs and as outputs. The stages identified in the context of this activity are the **data collection** phase -i.e. the processes involved in collecting datasets (qualitative and quantitative); data collection is usually followed by a distinct phase comprising the **preparation and/or the connection of datasets** -e.g., how to treat missing values and outliers, and/or the process of identifying individuals across the datasets. 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 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 is knowledge produced and verified.

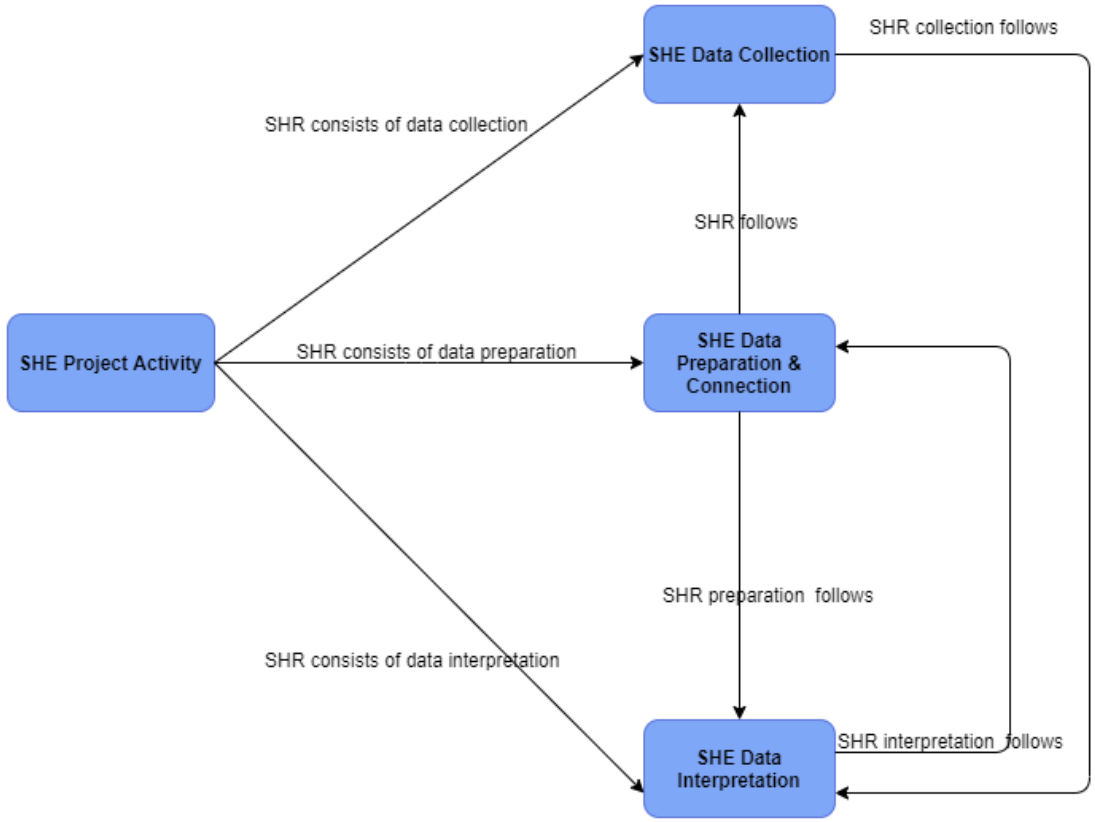

Figure 4: Parthood relations and temporal relations among SSHOCro activities

1.5 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).[2] In 2016, a group of researchers undertook the task of replicating 21 experimental studies that had been published in *Nature* and *Science* magazines, between 2010 and 2015, to determine the robustness of the reported results. These replications

[2] 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/

were conducted in the context of the Social Sciences Replication Project,[3] initiated by the Center for Open Science.[4]

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 (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 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 men's in a lab environment. The results extrapolate in policymaking.

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.[5] 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. **Figure 5** below represents the *Replication of Balafoutas and Sutter (2012)* at a macro-level. The *Replication of Balafoutas and Sutter (2012)* instantiates a Project Activity, which forms part of a larger Project Activity, undertaken by the group of researchers from the Innsbruck University division of the SSRP community. The SSRP, in its turn, is the overall Activity Project in the context of which the aforesaid Project Activities 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 SHE Project Activity, *SSRP_Replication of Balafoutas and Sutter (2012)*, can be further broken down to the series of discrete

[3] 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/

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

[5] 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/

sets of activities, each corresponding to a different stage in the research (collection, preparation-connection and interpretation).

The goal of the Project Activity 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 between the original and replication study, captured through *SHE Project Activity -SHR related: SHE Project Activity*. SHR related can be further specified –i.e. given a type –if there is adequate information to justify such a choice.

The rest of the properties connecting the *SSRP_Replication of Balafoutas and Sutter (2012)* inherited by E7 Activity –a superclass for SHE Project Activity.[6]

[6] Cf. Figure 1: CIDOC CRM E7 Activity and ICOM/CIDOC- CRM SIG (October 2019).

Figure 5: Replication of Balafoutas and 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 they put to the test. That dependence is manifold and relates to an effort for as little deviation as possible from the research protocols observed in the original studies, on behalf of the scholars conducting the replication. The processes and methods undertaken during data collection, data preparation-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 6**, below, aims for –namely to express the relation of datasets to the project activities that produced or deployed them in the course of a research project plus the relation between datasets across different projects.

In particular, this is the case with *SHE Dataset –SHR incorporates: SHE Dataset* (instantiated by the value pair "*Datasets created and analyzed in the context of Balafoutas & Sutter 2012*" and "*Balafoutas & Sutter (2012) - replicate data*", below).

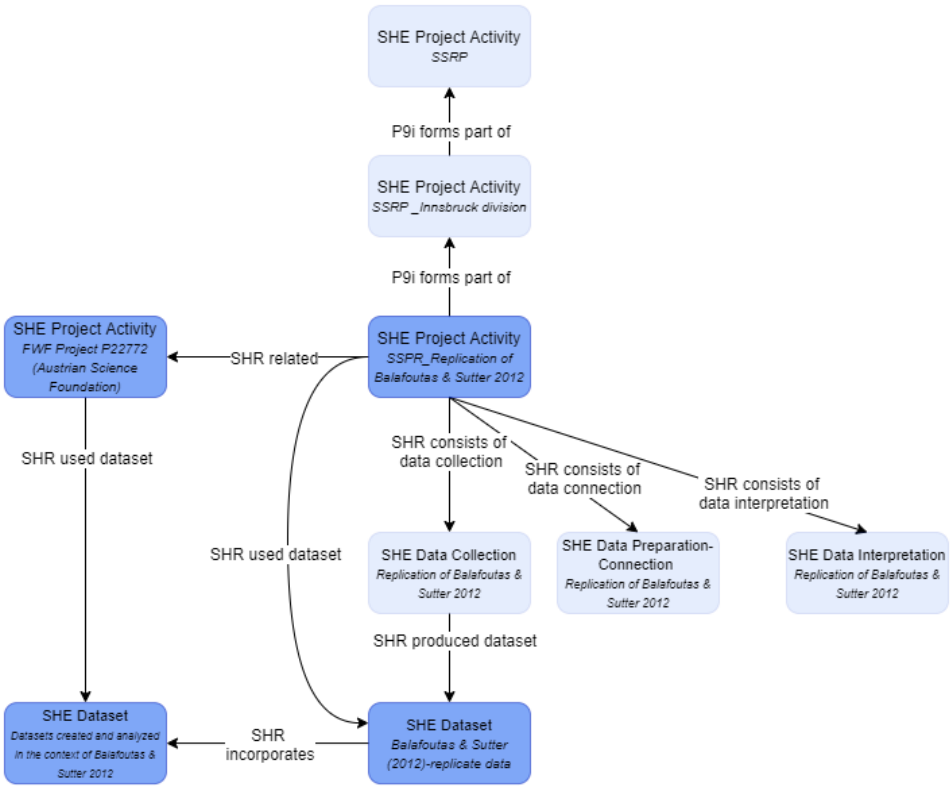

Figure 6: Data reuse in SSRP subprojects

Figures 7 through 10 illustrate the sets of activities undertaken during each of the distinct phases of the Project Activity *SSRP_Replication of Balafoutas and Sutter (2012)*. In particular, **Figure 7** illustrates the data collection process, **Figure 8** the stage of data preparation and connection, **Figure 9** the process of interpreting the collected data (post cleaning), and **Figure 10** 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 7 represents the data collection stage; it comprises data collection activity, linked to the time that it occurred –the place not being specified –to the actors that participated in it –and the capacity with which they

participated –the software used for collecting the data, the purpose the data collection served and its type – producing vs. finding and/or reusing.

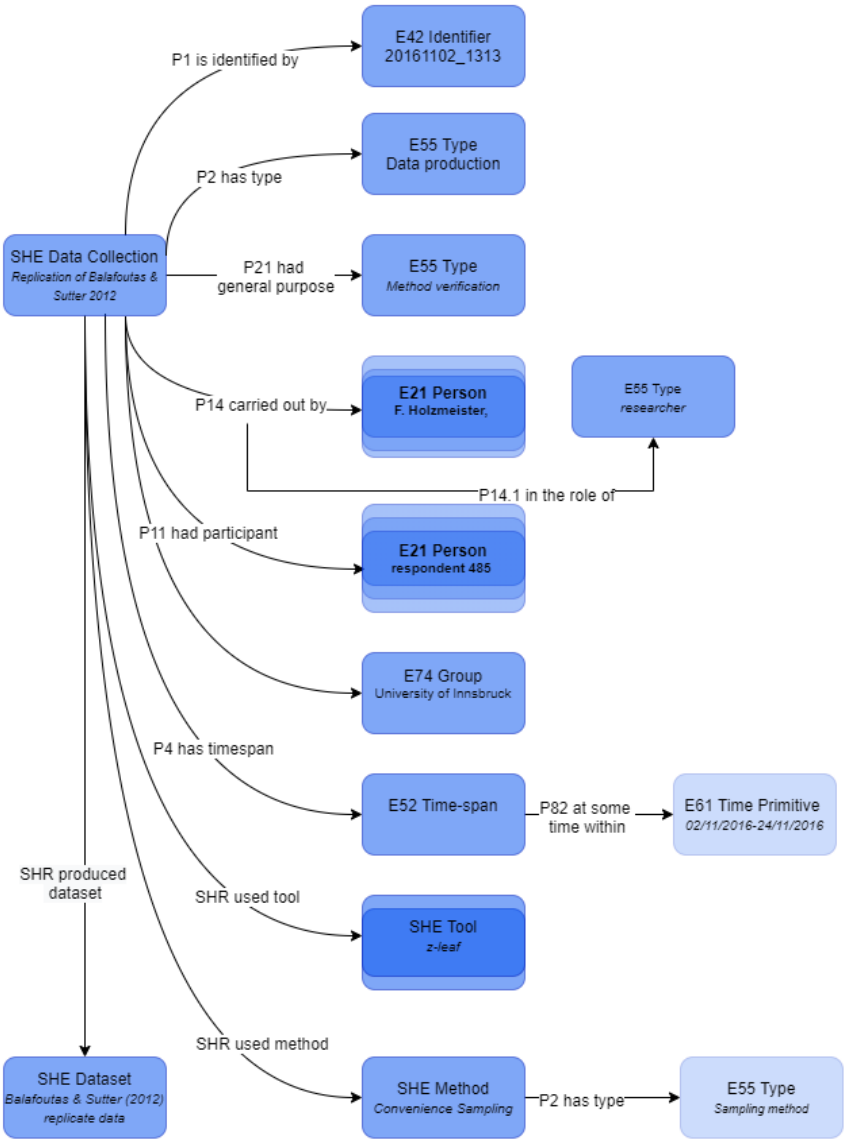

Figure 7: The data collection process

Figure 8 represents the data preparation-connection stage; it comprises a data preparation activity, linked to the time that it occurred –the place not being specified –to the actors that participated in it –and the capacity with which they participated –the software used for connecting the data, the methods deployed and their type. SHE Publication appears in the diagram above despite not being instantiated in the example, to demonstrate there can be publishing activities at all stages of the research project.

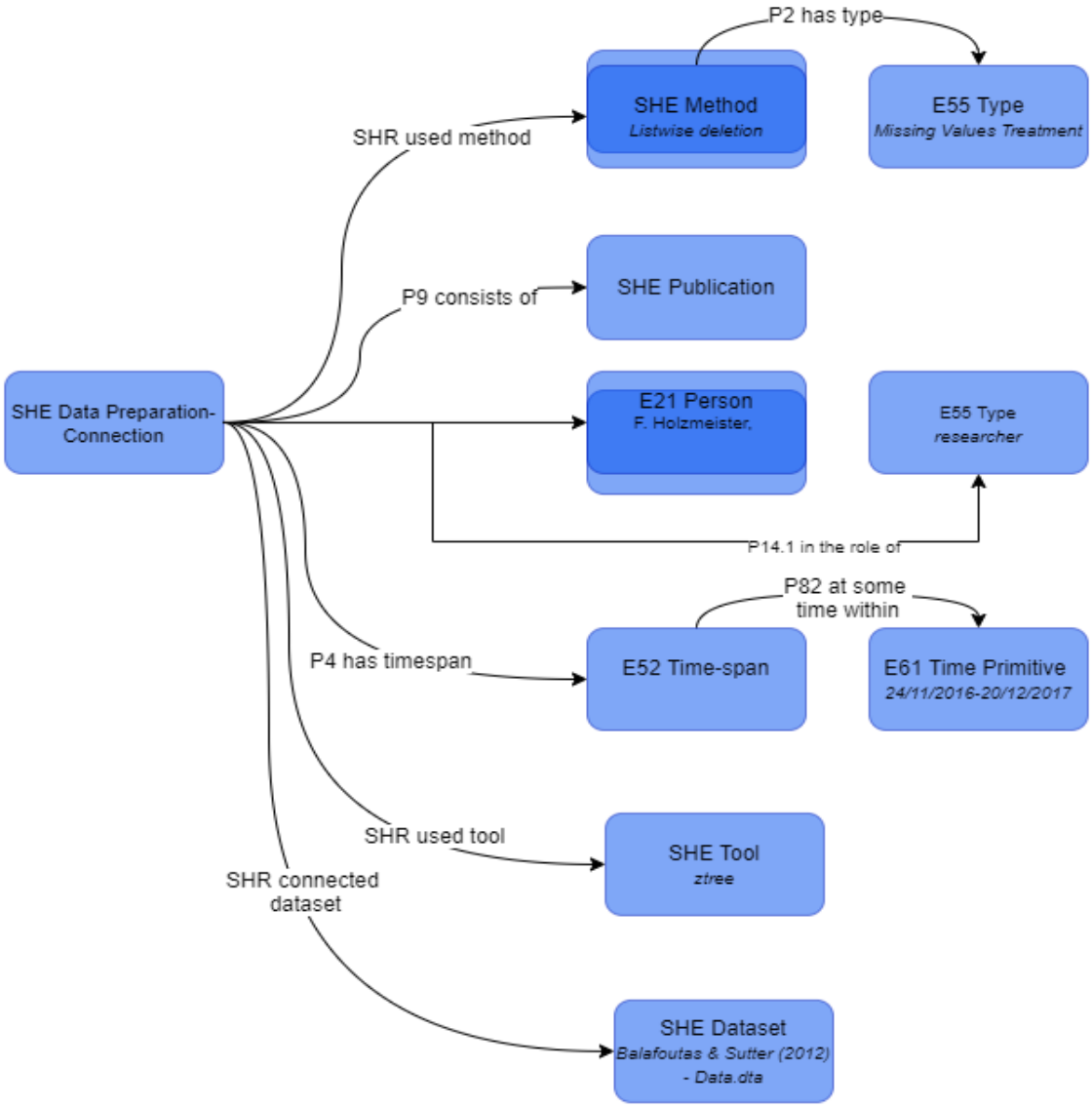

Figure 8: The data preparation-connection process (preprocessing phase)

Figure 9 represents the data interpretation; it comprises a data preparation activity, linked to the thing it was an interpretation of, the time during which it occurred, the actors that participated in it –and the capacity with which they participated –and the methods deployed for analyzing data and the software used in the analysis. The interpretation stage is further analyzed to its constituent parts –namely a set of activities of publishing data and replication reports/papers. Publishing activities are represented in greater detail in the following diagram.

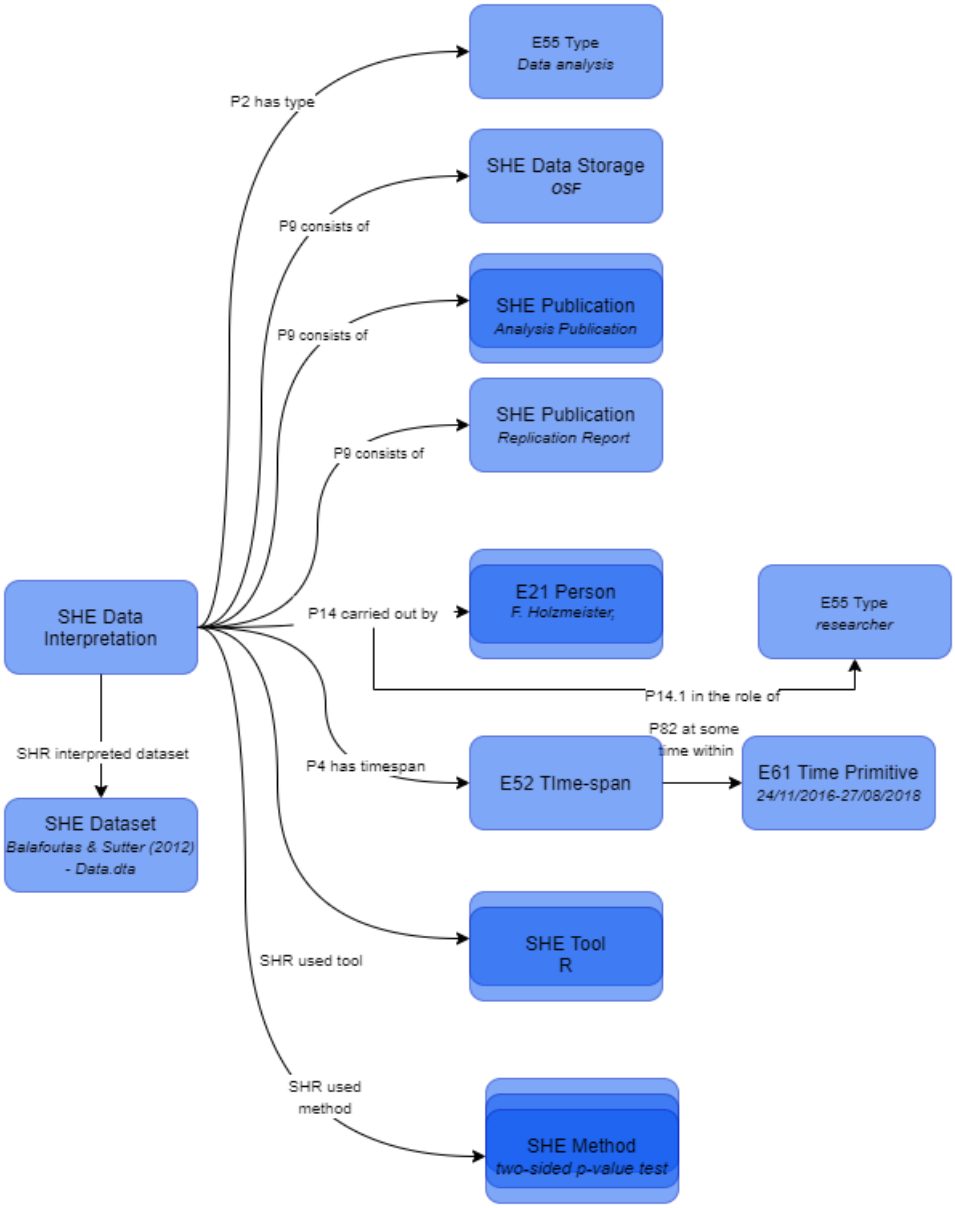

Figure 9: The data interpretation process

Figure 10 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 thing it published, the actor who performed it, the purpose of publishing it, the time it occurred and the kind of publication it involved.

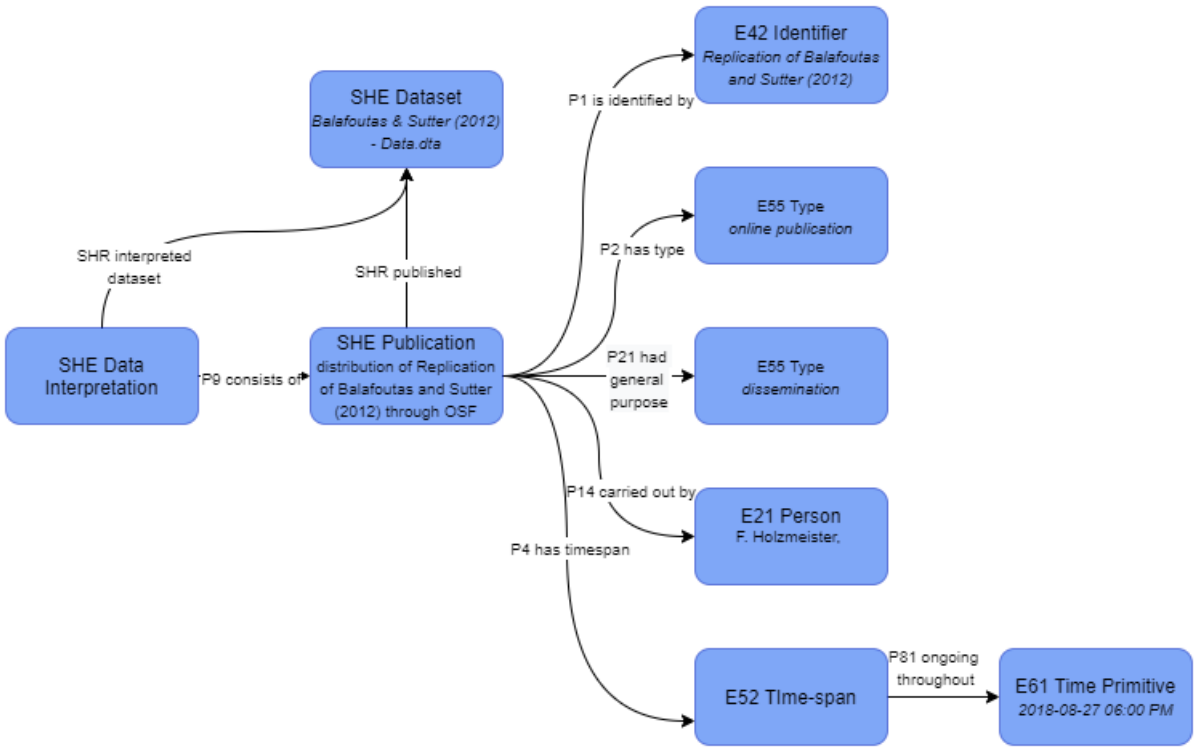

Figure 10: Publication of the results

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

—	E1 CRM Entity						
—	—	E7 Activity					
—	—	—	SHE Project Activity				
—	—	—	SHE Workflow Activity				
—	—	—	—	SHE Data Collection			
—	—	—	—	SHE Data Preparation & Connection			
—	—	—	—	SHE Data Interpretation			
—	—	—	—	SHE Publication			
—	—	—	—	SHE Data Storage			
—	—	E70 Thing					
—	—	—	SHE Tool				
—	—	—	E71 Human-Made Thing				
—	—	—	—	E28 Conceptual Object			
—	—	—	—	—	E90 Symbolic Object		
—	—	—	—	—	—	E73 Information Object	
—	—	—	—	—	—	—	SHE Method
—	—	—	—	—	—	—	SHE Dataset
—	—	—	—	—	E89 Propositional Object		
—	—	—	—	—	—	E73 Information Object	
—	—	—	—	—	—	—	SHE Method
—	—	—	—	—	—	—	SHE Dataset
—	—	—	—	—	E55 Type		
—	—	—	—	—	—	SHE Project Status	

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

PROPERTY NAME	ENTITY-DOMAIN	ENTITY RANGE
P9 consists of (forms part of)	E4 Period	E4 Period
— SHR consists of data collection (data collection part of)	SHE Project Activity	SHE Data Collection
— SHR consists of data preparation (data preparation part of)	SHE Project Activity	SHE Data Preparation Connection
— SHR consists of data interpretation (data interpretation part of)	SHE Project Activity	SHE Data Interpretation
P2 has type	E1 CRM Entity	E55 Type
— SHR has status	SHE Project Activity	SHE Project Status
P15 was influenced by (influenced)	E7 Activity	CRM Entity
— SHR based on (was basis for)	SHE Knowledge Workflow Activity	SHE Knowledge Workflow Activity
— P16 used specific object (was used for)	E7 Activity	E70 Thing
— — SHR had input (was input of)	SHE Knowledge Workflow Activity	SHE Dataset
— — — SHR used dataset for publication (was used by publication)	SHE Publication	SHE Dataset
— — SHR had output (was output of)	SHE Knowledge Workflow Activity	SHE Dataset
— — — SHR published	SHE Publication	SHE Dataset
— — — SHR produced dataset (dataset was produced by)	SHE Data Collection	SHE Dataset
— — SHR project used (was used by project)	SHE Project Activity	SHE Dataset
— — SHR used tool (was tool used for)	SHE Knowledge Workflow Activity	SHE Tool
— — SHR used method (was method used for)	SHE Knowledge Workflow Activity	SHE Method

P94 has created (was created by)		E65 Creation	E28 Conceptual Object
—	*SHR produced dataset (dataset was produced by)*	*SHE Data Collection*	*SHE Dataset*
P130 shows features of (features are also found on)		E70 Thing	E70 Thing
—	SHR has derivative (is derivative of)	SHE Dataset	SHE Dataset
P106 is composed of (forms part of)		E90 Symbolic Object	E90 Symbolic Object
—	SHR incorporates (is incorporated in)	SHE Dataset	
SHR related to		SHE Project Activity	SHE Project Activity
SHR offers (is offered by)		SHE Project Activity	SHE Service
SHR connected dataset (dataset was connected by)		SHE Data Connection Preparation	SHE Dataset
SHR interpreted dataset (dataset was interpreted by)		SHE Data Interpretation	SHE Dataset
SHR follows (is followed by)		SHE Data Preparation Connection	SHE Data Collection
SHR collection follows (is followed by collection)		SHE Data Collection	SHE Data Interpretation
SHR preparation follows (is followed by preparation)		SHE Data Preparation Connection	SHE Data Interpretation
SHR interpretation follows (is followed by interpretation)		SHE Data Interpretation	SHE Data Preparation Connection
SHR stored (was stored by)		SHE Data Storage	SHE Dataset

4. SSHOCro Class & Property Declaration

4.1 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

4.2 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

SHE 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 is 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:

- oral interviews
- Hand-written records
- Dataset of 50.000 definitions in Wikipedia
- Geospatial data
- Finland: FSD2883 Young People and Politics 2009: Essay Responses in Finnish
- ORCID Public Data File 2019 *(Published October 2019)*
- Balafoutas & Sutter (2012) - Data.dta

Properties: has derivative (is derivative of): SHE Dataset

incorporates (is incorporated in): SHE Dataset

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

SHE Knowledge Workflow Activity

Subclass of: E7 Activity

Superclass of: SHE Data Collection

SHE Data Preparation Connection

SHE Data Interpretation

SHE Publication

SHE Data Storage

Scope note: This class comprises the Knowledge Generation Process**:** represents the workflow of the processes used to produce specific datasets. It used as an entity which aggregates, generates specific data knowledge lifecycle activities characteristics.

Examples:

▪

Properties: SHR had output (was output of): SHE Dataset

SHR had input (was input of): SHE Dataset

SHR used tool (tool was used for): SHE Tool

SHR used method (method was used for): SHE Method

SHR based on (was basis for): SHE Knowledge Workflow Activity

Mapping to: MP Activity

SHE Project Activity

Subclass of: E7 Activity

Superclass of:

Scope note: This class comprises instances of collaborative (or sequential) activities undertaken over a period of time, carried out individually or collaboratively by an actor - the project team/consortium (person or group) and possibly involving research or design, that is carefully planned (usually by a project team) to achieve a particular aim or task and to produce a novel scientific contribution (under time or cost constraints).

Examples:

- Comparative Manifesto Project
- FWF Project P22772 (Austrian Science Foundation)
- The SEALIT[7] Project (ERC ID: 714437)

Properties: SHR consists of data collection (data collection part of): SHE Data Collection

[7] http://sealitproject.eu/?fbclid=IwAR20wxz6iHSscLK4mZGmCRBy-TxK3v0b5RHUBNVACvZixNICWyXlxu4rN8Y

SHR consists of data preparation (data preparation part of): SHE Data Preparation Connection

SHR consists of data interpretation (data interpretation part of): SHE Data Interpretation

SHR project used (was used by project): SHE Dataset

SHR related to: SHE Project Activity

SHR offers (is offered by): SHE Service

SHR has project status (is project status of): SHE Project Status

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

SHE Service

Subclass of: E7 Activity

Superclass of:

Scope note: This class comprises the offers declared by actor/s in order to perform an activity (that may result to a product) at the request of another actor/s that is going to gain benefit from this.

Examples:

- CESSDA service
- SND (Swedish National Data Service)

Properties:

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

SHE Data Collection

Subclass of: SHE Knowledge Workflow Activity

E65 Creation

Superclass of:

Scope note: This class comprises the evidence collection, the gathering of objects of common criteria (qualitative and quantitative) using different methods and policies such as surveys types, research methods or by direct observation.

Examples:

- Data collection for replication by Balafoutas & Sutter (2012)

Properties: SHR produced dataset (dataset was produced by): SHE Dataset

SHR collection follows (is followed by collection): SHE Data Interpretation

Mapping to:

SHE Data Preparation Connection

Subclass of: SHE 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[8] records in SEALIT project.

Properties: SHR connected dataset (was connected by): SHE Dataset

SHR follows (is followed by): SHE Data Collection

SHR preparation follows (is followed by preparation): SHE Data Interpretation

Mapping to:

SHE Data Interpretation

Subclass of: SHE Knowledge Workflow Activity

Superclass of:

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:

[8] http://139.91.183.60:8181/FastCat/index.html
http://www.sealitproject.eu/

- Hypothesis testing: Replication of experiments is possible and yields reliable results in social sciences
- The replicability of Balafoutas & Sutter (2012) by F. Holzmeister, J.Huber, M. kirchler, J.Rose supports the replicability in social sciences hypothesis.
- The deduction (made by A.Evans) on the existence of a pottery workshop in Knossos in the early years of the 20th century AD.

Properties: SHR interpreted dataset (dataset was interpreted by): SHE Dataset

SHR interpretation follows (is followed by interpretation): SHE Data Preparation Connection

Mapping to:

SHE Publication

Subclass of: SHE Knowledge Workflow Activity

Superclass of:

Scope note: This class comprises the data sharing (including any conditions or restrictions on that) and publication activities in order to maximize the data's discoverability.

Examples:

- "Replication of Balafoutas and Sutter (2012)" (Online publication)

Properties: SHR used dataset for publication (was used by publication): SHE Dataset

SHR published (was published by): SHE Dataset

Mapping to: CERIF Publication

SHE Data Storage

Subclass of: SHE Knowledge Workflow Activity

Superclass of:

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

Examples:

- Balafoutas & Sutter (2012) - Data.dta Storage using Innsbruck University server

Properties: SHR stored (was stored by): SHE Dataset

Mapping to:

SHE Tool

Subclass of: E70 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 [9]
- 3M Editor[10]
- Box
- SPSS[11]

Properties:

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

SHE Method

Subclass of: E73 Information Object

Superclass of:

Scope note: This class comprises the specifications, 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
- Listwise deletion

Properties:

Mapping to: SO Method

SHE Project Status

Subclass of: E55 Type

[9] STATA: https://www.stata.com/

[10] #3M Editor: http://web-new.ics.forth.gr/isl/3m-editor

[11] SPSS: https://www.ibm.com/analytics/spss-statistics-software

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

6. SSHOCro Properties

SHR consists of data collection (data collection part of)

Domain: SHE Project Activity

Range: SHE 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 analyzed 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) Project *consists of data collection* experiment data collection from 02/11/2016 to 24/11/2016.

Mapping to:

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

SHR consists of data preparation (data preparation part of)

Domain: SHE Project Activity

Range: SHE Data Preparation 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 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) Project *consists of data preparation* data entry (on 2017-12-20).

Mapping to:

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

SHR consists of data interpretation (data interpretation part of)

Domain: SHE Project Activity

Range: SHE 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:

- SSRP – Innsbruck: Replication of Balafoutas & Sutter (2012) Project *consists of data interpretation* hypothesis testing on Balafoutas & Sutter (2012) - Data.dta v1.0 from 24/11/2016 to 27/8/2018.

Mapping to: SO Activity: SO part of: SO Activity

MP Activity: composed of: MP Activity

SHR had input (was input of)

Domain: SHE Knowledge Workflow Activity

Range: SHE Dataset

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

Superproperty of: SHE Publication. SHR used dataset for publication (was used by publication): SHE Dataset

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

Examples:

Mapping to:

MP Activity: inputEntity: MP Entity

SHR had output (was output of)

Domain: SHE Knowledge Workflow Activity

Range: SHE Dataset

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

Superproperty of: SHE Publication. SHR published (was published by): SHE Dataset

SHE Data Collection. SHR produced dataset (dataset was produced by): SHE Dataset

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

Examples:

Mapping to:

MP Activity: outputEntity: MP Entity

SHR project used (was used by project)

Domain: SHE Project Activity

Range: SHE 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 SHE Project Activity.

Examples: Replication of Balafoutas & Sutter (2012) Project *used* Balafoutas & Sutter (2012) - Data.dta.

Mapping to:

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

SHR related to

Domain: SHE Project Activity

Range: SHE Project Activity

Subproperty of:

Superproperty of:

Scope note: This property describes, in general, the existence of a relation between SHE Project Activities, without specifying the kind of relation (e.g a project is related to another project because it may use methods or data or experiments that were conducted in the context of that project).

Examples:

- Social Sciences Replication Project [SSRP] (replicate) *related to* FWF Project P22772 (Austrian Science Foundation).

Mapping to:

- PARTHENOS
  PE35 Project. PP43 supports project activity (is project activity supported by): CRM E7 Activity,
- CERIF Project relationship
- MP Entity: related: MP Entity

SHR offers (is offered by)

Domain: SHE Project Activity

Range: SHE Service

Subproperty of:

Superproperty of:

Scope note: This property associates an instance of a project with the instance of the service(s) it presently offers.

Examples:

- SSRP –Innsbruck Division *offers* the Social Sciences Replication on OSF.
- The Ariadne Project *offers* the Ariadne catalogue.

Mapping to:

- PARTHENOS PP1
  PE26 RI Project. PP1 currently offers: PE1 Service

- CERIF Project Service relationship
- MP Activity:usedTools: MP Service

SHR has project status (is project status of)

Domain: SHE Project Activity

Range: SHE 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* past.

Mapping to: CERIF Project Class relationship

MP Activity: property:MP Property:concept:MP Concept

SHR produced dataset (dataset was produced by)

Domain: SHE Data Collection

Range: SHE Dataset

Subproperty of: SHE Knowledge Workflow Activity. SHR had output (was output of): SHE Dataset

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:

- Experiment data collection 02/11/2016-24/11/2016 *produced dataset* Balafoutas & Sutter (2012) - Data.dta v1.0.

Mapping to: SO produces

- Scholarly Ontology:
  SO Activity. SO produces: SO InformationResource

SHR connected dataset (was connected by)

Domain: SHE Data Preparation Connection

Range: SHE Dataset

Subproperty of:

Superproperty of:

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

Examples:

- Data-entry from 24/11/2016 to 20/12/201 *connected dataset* Balafoutas & Sutter (2012) - Data.dta.

Mapping to:

SHR interpreted dataset (was interpreted by)

Domain: SHE Data Interpretation

Range: SHE Dataset

Subproperty of:

Superproperty of:

Scope note: This property identifies a dataset that was the result of an analysis/interpretation activity.

Examples:

- Replicability hypothesis testing by F. Holzmeister, J.Huber, M. kirchler, J.Rose *interpreted dataset* Balafoutas & Sutter (2012) - Data.dta v1.0.

Mapping to:

SHR follows (is followed by)

Domain: SHE Data Preparation Connection

Range: SHE Data Collection

Subproperty of:

Superproperty of:

Scope note: This property describes the sequence relation between two successive activities. This property indicates that one instance of data preparation 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:

- Data Entry (replicate) *follows* Data Collection for replication by Balafoutas & Sutter (2012).

Mapping to:

- Scholarly Ontology:
  SO Activity. SO follows: SO Activity
- MP Activity: follows: MP Activity

SHR collection follows (is followed by collection)

Domain: SHE Data Collection

Range: SHE Data Interpretation

Subproperty of:

Superproperty of:

Scope note: This property describes the sequence relation between two activities. 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, rationalization, and validation.

Examples:

Mapping to:

- Scholarly Ontology:
  SO Activity. SO follows: SO Activity
- MP Activity: follows: MP Activity

SHR preparation follows (is followed by preparation)

Domain: SHE Data Preparation Connection

Range: SHE Data Interpretation

Subproperty of:

Superproperty of:

Scope note: This property describes the sequence relation between two activities. This property indicates that one instance of data preparation 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, rationalization, and validation.

Examples:

Mapping to:

- Scholarly Ontology:
  SO Activity. SO follows: SO Activity
- MP Activity: follows: MP Activity

SHR interpretation follows (is followed by interpretation)

Domain: SHE Data Interpretation

Range: SHE Data Preparation Connection

Subproperty of:

Superproperty of:

Scope note: This property describes the sequence relation between two successive activities. This property indicates that one instance of data interpretation activity follows an instance of data preparation 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:

- Replicability hypothesis testing on Balafoutas & Sutter (2012) - Data.dta v1.0 *interpretation follows* replication data entry.

Mapping to:

- Scholarly Ontology:
  SO Activity. SO follows: SO Activity
- MP Activity: follows: MP Activity

SHR used dataset for publication (was used by publication)

Domain: SHE Publication

Range: SHE Dataset

Subproperty of: SHE Knowledge Workflow Activity. SHR had input (was input of): SHE Dataset

Superproperty of:

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

Examples:

- Replication Report *used dataset for publication* Balafoutas & Sutter (2012) - Data.dta v1.0.

Mapping to:

- Scholarly Ontology:
  SO Activity. SO uses: SO InformationResource

SHR published (was published by)

Domain: SHE Publication

Range: SHE Dataset

Subproperty of: SHE Knowledge Workflow Activity. SHR had output (was output of): SHE Dataset

Superproperty of:

Scope note: This property associates an instance of a dataset with an instance of a publication that published it.

Examples:

- Publication by Camerer, C.F., Dreber, A., Holzmeister, F. et al. (2018) *published* Complete Repository (.zip)

Mapping to:

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

SHR stored (was stored by)

Domain: SHE Data Storage

Range: SHE Dataset

Subproperty of:

Superproperty of:

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

Examples:

- "Data storage using Innsbruck University server" *stored* " Balafoutas & Sutter (2012) - Data.dta".

Mapping to:

SHR has derivative (is derivative of)

Domain: SHE Dataset

Range: SHE Dataset

Subproperty of: E70 Thing. P130 shows features of (features are also found on): E70 Thing

Superproperty of:

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:

- "Balafoutas & Sutter (2012) Data" [OR] has derivative "Balafoutas & Sutter (2012) Data.dta v1.0"

Mapping to:

MP Dataset: related: MP Dataset

SHR incorporates (is incorporated in)

Domain: SHE Dataset

Range: SHE Dataset

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

- "Balafoutas & Sutter (2012) - Data.dta v1.0" *incorporates* "Balafoutas & Sutter (2012) Data" [OR].

Mapping to:

MP Dataset: related: MP Dataset

SHR used tool (tool was used for)

Domain: SHE Knowledge Workflow Activity

Range: SHE 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:

- Experiment data collection *used tool* z-tree.

Mapping to:

- Scholarly Ontology:
  SO Activity. SO usesTool: SO Tool
- MP Activity: usedTools: MP Tool

SHR used method (method was used for)

Domain: SHE Knowledge Workflow Activity

Range: SHE 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.

Mapping to:

- Scholarly Ontology:
  SO Activity. SO employs: SO Method

SHR based on (was basis for)

Domain: SHE Knowledge Workflow Activity

Range: SHE 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:

- Replicate experiment data collection was *based on* original experiment data collection.

Mapping to:

MP Activity: related: MP Activity

7. Referred CRM Classes & Properties

Since the SSHOCro ontology refers to and reuses, wherever appropriate, large parts of ISO21127, the CIDOC Conceptual Reference Model, this section provides a comprehensive list of all constructs used from ISO21127, together with their definitions following version 6.2.7 maintained by CIDOC. Use in this context includes: reference as immediate superclass, superproperty or element of a path expression in a mapping statement.

7.1 Referred CRM Classes

This section contains the complete definitions of the classes of the CIDOC CRM Conceptual Reference Model version 6.2.7 referred to by SSHOCro ontology.

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, 1987)
- 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

E55 Type

Subclass of: E28 Conceptual Object

Superclass of: E56 Language

E57 Material

E58 Measurement Unit

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)*: E55Type. Such hierarchies may be extended with additional properties.

Examples:

- weight, length, depth [types of E54]
- portrait, sketch, animation [types of E36]
- French, English, German [E56]
- excellent, good, poor [types of E3]
- Ford Model T, chop stick [types of E22]
- cave, doline, scratch [types of E26]
- poem, short story [types of E33]
- wedding, earthquake, skirmish [types 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

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)
- the bottle of milk in my refrigerator (E22)
- the plan of the Straßburger Münster (French: *Cathédrale Notre-Dame de Strasbourg*) (E29)
- 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 (E27) (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)

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 a 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)
- E. A. Poe's "The Raven" (Poe, 1869)
- the movie "The Seven Samurai" by Akira Kurosawa (Mellen, 2002)
- 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:

7.2 Referred CRM Properties

This section contains the complete definitions of the properties of the CIDOC CRM Conceptual Reference Model version 6.2.7 referred to by SSHOCro ontology.

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

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. Sub typing obviously requires consistency between the meaning of the terms assigned and the more general intent of the CIDOC CRM entity in question.

Examples:

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

In First Order Logic:

P2(x,y) ⊃ E1(x)

P2(x,y) ⊃ E55(y)

P9 consists of (forms part of)

Domain: E4 Period

Range: E4 Period
Subproperty of: E92 Spacetime Volume. P10i contains (falls within):E92 Spacetime Volume

Quantification: one to many, (0,n:0,1)

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.

Examples:

- Cretan Bronze Age (E4) *consists of* Middle Minoan (E4)

In First Order Logic:

P9(x,y) ⊃ E4(x)

P9(x,y) ⊃ E4(y)

P9(x,y) ⊃ P10(y,x)

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)

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 this scope note (E7) *used specific object* Nicholas Crofts' computer (E22) *mode of use* Typing Tool; Storage Medium (E55)
- 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

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.

Examples:

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

In First Order Logic:

P106(x,y) ⊃ E90(x)

P106(x,y) ⊃ E90(y)

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.

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

8. References

Balafoutas, L. & Sutter, M. (2012). *Affirmative action policies promote women and do not harm efficiency in the laboratory.* Science 335, 579–582

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

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.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/

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

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., 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), 34 pages. DOI: https://doi.org/10.1145/1921614.1921615

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

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.

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/

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

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.

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

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

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

9. List of Figures