Semper-Edition

Edition team: Michael Gnehm, Carmen aus der Au, Tanja Kevic, Dieter Weidmann
by Elena Chestnova and Swen Vermeul

Introduction

The toolbox of the humanities scholar has been undergoing a digital expansion ever since the late 1980s. On perhaps the most basic level, the character of this expansion has been defined by the explosion in digital and digitized sources – primary source texts of all varieties (whether born-digital or not) made available in a digital form. Alongside this, a number of digital methods for working with text have emerged in recent decades. Many arguments in favour of digitisation – of both source and method – are indisputable: widening access, processing large corpora, combining datasets that cannot be combined by analogue approaches to generate new insight. But caution has also been advised when embracing new

approaches in the digital humanities: digital analysis often falls short of representing, or, in the worst cases, actively distorts lived experience and reproduces bias instead of exposing it. Digital tools and sources call for a critical attitude that would see them integrated among the humanists' toolbox as a complement and extension, sometimes radical, to theories and approaches of 'analogue' pedigree. Such methods and resources would support scholars in the study of their materials in ways that are inaccessible in the analogue while giving them the flexibility to integrate different types of analysis. In addition, recent scholarship has argued for extensive application of analytical critique to digital methods and approaches both with the view to comprehend their nature and implications, and to interrogate the different types of cultural baggage and bias they
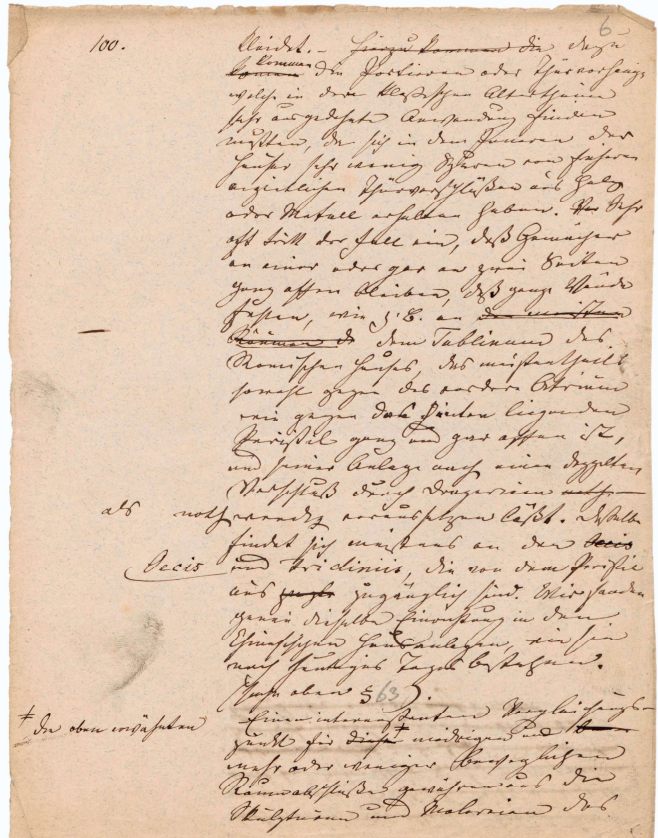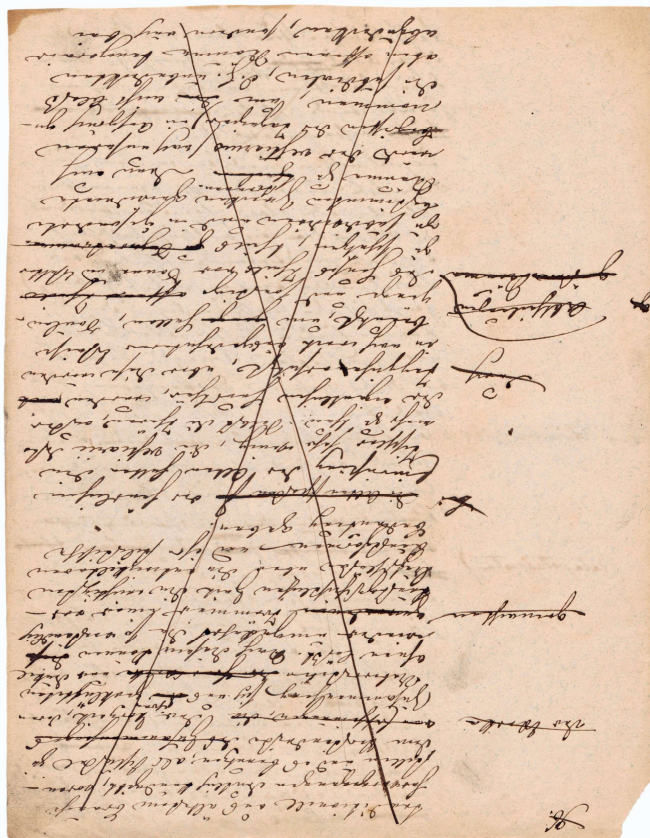

Figure 1. Gottfried Semper. Draft manuscript for Style vol. 1. gta Archives 20-Ms-218, fols. 5v-6r. Copyright: gta Archives / ETH Zurich.

carry.[1] Transparency, as one of the main ideals of this paradigm, would enable a detailed analysis and critique of how a particular source or method takes place or comes to be. This entails not only publishing assumptions behind a particular project of digitization, but also giving an account of its workflow and detailed descriptions of what happens in each of the technical processes applied. In this paper we present a brief account of a novel digital edition of Gottfried Semper's *Style* that aims to equip the reader with first information to begin such an analysis.

"Gottfried Semper: *Style*. Critical and commented edition" is a collaborative project of the ETH Zurich (responsible: Philip Ursprung) and the Università della Svizzera italiana (responsible: Sonja Hildebrand), funded by the Swiss National Science Foundation (SNSF). It is expected to last up to twelve years and it is coming up to the end of year four – a milestone in its intended three-part chronology. The project aims to digitize, transcribe and make available to the public sources associated with the magnum opus of Gottfried Semper – a major German architect and theorist of art and architecture. This book, which appeared in two volumes in 1860 and 1863 under the title *Der Stil in den technischen und tektonischen Künsten, oder Praktische Ästhetik*, remains widely read not only by art and architectural historians and theoreticians but also by practicing architects and designers, among others. Its evolution spans several decades of drafts which have so far been only partially transcribed and published.[2]

The Edition's corpus is highly heterogeneous, including autograph manuscripts in German Kurrentschrift (containing fragments in Latin and Greek scripts) with varying degrees of authorial intervention (fig. 1), calligraphic manuscript copies in Kurrentschrift (fig. 2), print proofs with hand-written corrections (fig. 3), and different versions of the printed text. The corrections in the text, even those undertaken in the last stages of editing, and remarks included in its margins often contain a wealth of historical information about the book. So, for example, page 514 of the first volume of edition 1 was commented by Semper in the proof stage to reflect the death of the German art historian Franz Kugler – one of his harshest critics, whose remarks are addressed in the final pages of volume 1 of *Style*. Its publication was delayed and Kugler passed away in the meantime, which led Semper to consider removing the relevant section. The text was finally left untouched with Semper motivating his decision as follows (fig. 3):

"Diese Schlussbemerkungen sind noch bei Lebzeiten des Prof. Kugler niedergeschrieben worden, wie leicht ersichtlich ist. – Es
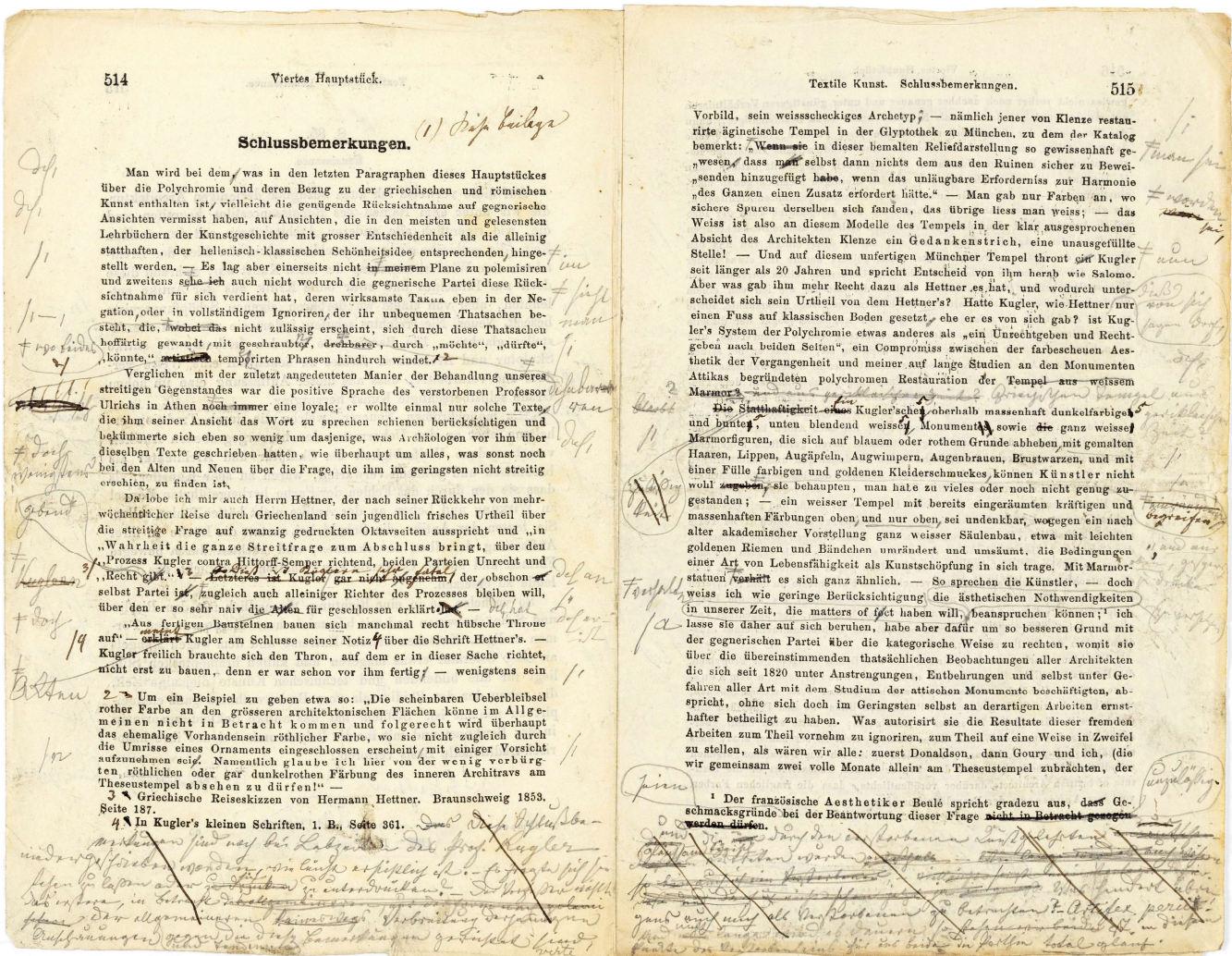

Figure 3. Gottfried Semper. Print proofs for Style vol. I, pp. 514-515. gta Archives 20-Ms-252-82. Copyright: gta Archives / ETH Zurich.

fragte sich sie stehen zu lassen oder sie zu unterdrücken? – Der Verfasser wählte das erstere, in Betracht der allgemeinen Verbreitung derjenigen Anschauungen und Tendenzen gegen die diese Bemerkungen gerichtet sind, und die durch den verstorbenen Kunstgelehrten nur gleichsam typisch vertreten wurden. Was hindert übrigens auch mich als Verstorbenen zu betrachten? – Artifex periit! Und nicht lange wird es dauern so ist in diesem Punkte des Verstorbenseins für uns beide die Parthie total gleich."[3]

This comment is particularly important in light of the relationship between the ideas of the two men. Kugler ridiculed Semper's proposition that architecture of ancient Greece had been painted, rather than white, as was assumed previously. Although Semper had many authorities on his side, Kugler's remarks stung, leading him to return to the issue in several publications.[4] The decision to retain his critical remarks in the conclusion of the first volume of *Style* allowed Semper to have the last word in the dispute, reflecting its importance, in his eyes, for his theoretical enterprise.

Workflow

The field of digital editing, dynamic since its establishment in the 1980s, currently finds itself at a crossroads. On the one hand, TEI (Text Encoding Initiative) is considered the de facto standard for the digitization of texts in all areas of the humanities. On the other hand, its limitations in the capture of complex text corpora are well-recognized.

XML (Extensible Markup Language) is a markup language that defines a set of rules for encoding documents in a format that is readable by humans and machines with relatively simple technical means. The standard of TEI defines how XML is used for the representation of texts in the scientific field. Since XML itself is basically just text, its preservation is fundamentally simple.[5] Published TEI XML transcripts enable an expert reader to acquire a deeper understanding of the methods and assumptions behind a digital edition.[6] However, the hierarchical structure of XML often makes it difficult to represent real text phenomena, e.g. overlapping tags are not allowed. This problem is solved either by using so called stand-off markup (as opposed to inline markup) or by applying empty elements or milestones. But these techniques tend to compromise the advantages of XML – the complexity of markup increases and the texts become very difficult for humans to read, while the indexing required for the stand-off solution makes the process of annotation too rigid. The limited potential of TEI XML to represent complex text and text-data contexts has led to the proposal of several alternative approaches.[7] These approaches, however, continue to be considered cutting-edge, and while their future viability is an important part of their brief, TEI XML remains a mainstay of best practice in creating digital editions. This issue has been addressed in

recent scholarship, and calls have been made to develop Janus-faced solutions suspended between best practice and innovation, which on the one hand retain the established practices and standards, but on the other hand open up the potential of text resources for future research and use.[8] The Semper-Edition follows this path, in that it maintains TEI XML as a key part of its workflow, while final versions of transcriptions are converted and stored in the labelled property graph database Neo4j.

The heterogeneous character of the Edition's corpus demands a differentiated workflow with a highly standardized output suitable for automated import into the graph database. The relatively high volume of manuscripts to be transcribed calls for the application of automated hand-written text recognition (HTR) and optical character recognition (OCR). However, the possible extent of this and, accordingly, the respective procedure, vary between different types of sources, producing a complex workflow (fig. 4).

Efficient representation of graphical information is achieved in the TEI XML transcriptions of the Semper-Edition through the use of the elements "facsimile" and "zone", which denote the position of individual text fields on the facsimile scan of the manuscript page with coordinates. This information is generated with the help of tools provided by the Transkribus[9] software, which also enables a degree of automation for layout recognition. This forms the first step in the processing of all sources of the edition. After that, the processing follows a varied path depending on the type of source. Manuscripts with a minimal degree of intervention, calligraphically written by copyists in German Kurrentschrift, are processed with an HTR model. Print text is processed with OCR. These transcriptions are then read and corrected in Transkribus, exported to TEI XML, converted with XSLT and other tools available in Oxygen XML Editor and enriched with further information, for example, attributes that denote the reading order of paragraphs. Where print text contains handwritten notes and corrections, these are likewise transcribed in Transkribus. In the case of the complex autograph manuscripts, only the layout is generated and adjusted in Transkribus. It is then exported by generating a TEI XML file that contains only the page structure. The whole transcription is in this case undertaken directly in XML.

Semper's texts are extremely rich in references to other texts and people, artefacts, places and organizations. They also contain many instances of terms whose use could be productively interrogated, such as names used to denote various groupings of people like "Chaldeans" or "Assyrians," or conceptual terms specifically associated with Semper's theoretical corpus, such as "Bekleidung." The importance of these references has been highlighted in the calls to develop a "way of tackling digital editing not only as a more efficient means than a paper edition to report text genesis but also as a possible recreating of the whole intellectual, cultural, and contemporary historical
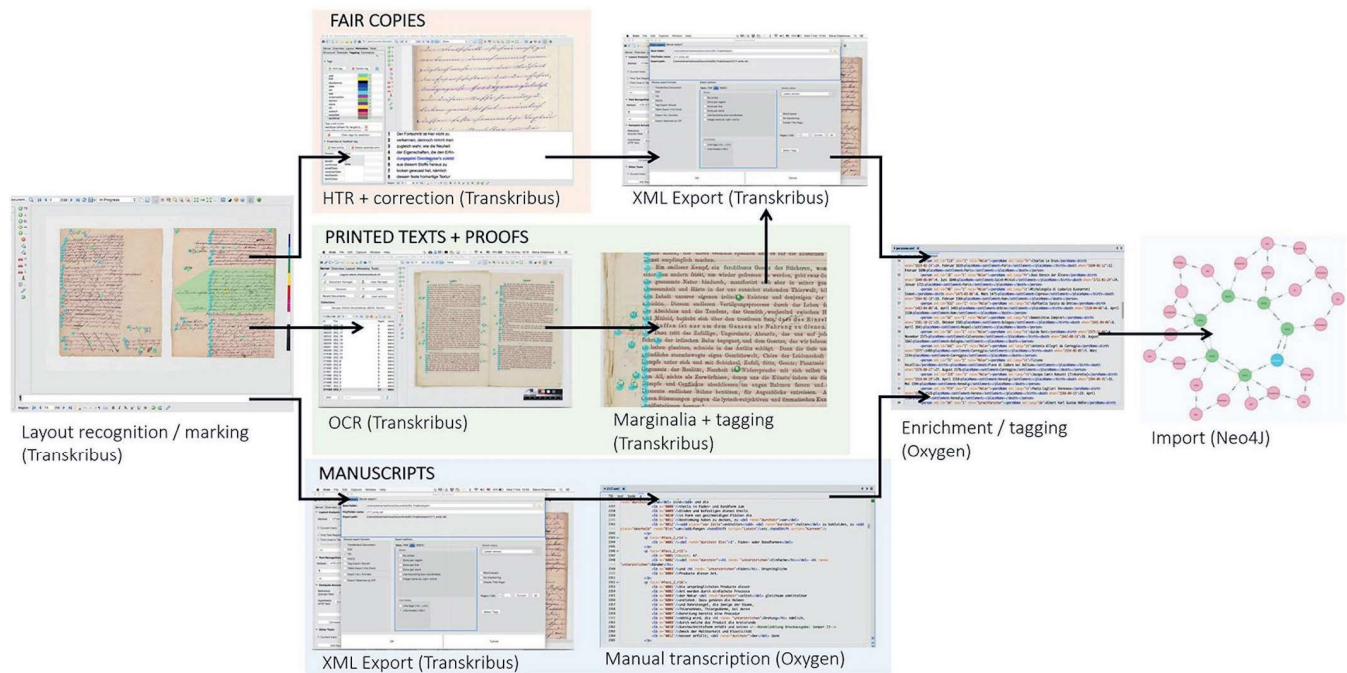
FAIR COPIES

HTR + correction (Transkribus)

XML Export (Transkribus)

PRINTED TEXTS + PROOFS

Layout recognition / marking
(Transkribus)

OCR (Transkribus)

Marginalia + tagging
(Transkribus)

Enrichment / tagging
(Oxygen)

Import (Neo4J)

MANUSCRIPTS

XML Export (Transkribus)

Manual transcription (Oxygen)

Figure 4.   Workflow of the Semper-Edition.

universe of its birth."[10] In order to enable the Edition and its readers to undertake such reconstructions in the future, named entities (persons, places, bibliographic references, organizations, peoples, artefacts, and concepts) are tagged in all XML TEI files and are linked by the means of unique identifiers to the relevant index of the Edition.

The importance of Graph Databases

Semper-Edition uses a labelled property graph database (Neo4j, version 4) to store its data. In graph databases, every piece of information is stored in a *node*. A node is tagged with one or several *labels*, such as Person, Token, Paragraph, Zone. In addition, every node contains properties, key-value pairs that contain the actual information. Nodes can have *relations* to any number of other nodes. A relation is always directed (A)-[:knows]→(B), can have only one label but many properties. A number of nodes that have relations to other nodes are called a *graph*. Relations act like indexes; they are our means to quickly find the shortest or best path through our graph.

In contrast, traditional relational databases (RDB) store their data in tables. All similar pieces of information are stored in the same table. A relation between different pieces of information is stored in an additional table that contains the primary keys of the two tables. Despite their name, relational databases show serious performance problems when a lot of relations are involved (often called "join bomb").

A graph database is much quicker to establish such queries because a join between two nodes is as efficient as running through an index: no lookup of data is involved. The result of a search query in a traditional RDB always returns a table. In a graph database, the search might start at a given node and then tries to find a path to possible ends. The result of such a query is often a number of paths. In other words, the result of a graph database query can be a complex data structure.

These features suggest that graph databases are extremely well suited for the representation and preservation of complex, linkage-rich data. Test reports in latest scholarly discussion have evaluated Neo4j compared to other databases, and demonstrated its high performance and capacity for preserving the complexity of data.[11] In addition, Neo4j allows easy linking of data under the Linked Data paradigm (parts of the graph can conform to the Resource Description Framework RDF and be queried by SPARQL).

The digital humanities are making increasing use of property graphs to store, retrieve and process complex data and to represent text as graph (TAG). TAG applications have become quite common in the fields of digital editing and digital text analysis in recent years.[12]

Text as Graph (TAG)

The ability of property graph databases to store links between data points has prompted the use of this platform for innovative ways of presenting edited text.[13] As part of the

Text-As-Graph (TAG) paradigm, text is stored as chains of nodes and relations (fig. 5). At the simplest level, nodes are used to denote so-called tokens, which represent words, parts of words or punctuation marks. Whitespace that separates two tokens is stored as a property of the node. We store a number of properties for every token, such as the currently used script (Kurrent or Latin), the part-of-speech (POS) tag of the Spacy tokenizer (e.g. verb, noun, proposition, punctuation) and the current paragraph within the document.

These token-nodes are then linked with other tokens with a relation to indicate the direction of the textflow, e.g. (This)-[:NEXT]→(is)-[:NEXT]→(the)-[:NEXT]→(flow). Other types of nodes mark the positions of line breaks.

paragraphs and pages. A paragraph node usually has a relation to the first line-break, which then points to the first token within that line. Further relations connect tokens with relevant data, e.g. the word "Assyrian" is connected to the corresponding entry in the register.

Import of TEI XML into the Graph Database

TEI XML is imported into Neo4j in the Semper-Edition by means of a Python script developed by the Scientific Information Services of ETH Zurich and published in the Python repository (https://pypi.org/project/tei2neo/). The script is built of several parts:

Figure 5. Simplified representation of a transcription of a short page of manuscript text (gta 20-Ms-215 fol. 5r) in the graph database of the Semper-Edition. Facsimile shown on the right. Copyright: gta Archives / ETH Zurich.

blue = paragraphs
yellow = linebreaks
rose = words / tokens
red = deleted words
green = added word

– an XML parser (lxml, Beautiful Soup) splits the TEI XML document into data units, e.g. header information, page breaks, paragraphs, line breaks and text
– the text is then split into tokens (word, parts of words, punctuation, symbols, etc.) using Spacy tokenizer. The Spacy algorithm also enhances the tokens with additional syntactic information.
– certain XML elements affect the appearance of their contained elements (e.g. del, add, highlight)
– other XML elements affect the appearance of all follo-

wing elements until a marker is found, such as delSpan or handShift (change of writing style)
- all XML elements and all tokens are stored as nodes with relations, using the py2neo library
- hyphened words are concatenated again and stored as an alternative node, with an alternative text path
- line breaks, paragraphs and pages are connected
- paragraphs that span over two pages are connected
- paragraphs are connected to their corresponding zones (coordinates on the facsimile)

The detailed data from persons, places, terms, artefacts, organizations and bibliographical references is also imported into the graph database. Every node can refer to one or many register entries. Because of the graph database, we get the back-references for free.

Retrieving Information from the Graph Database

In contrast to other projects which consider variations of the text as a property of the text chain and where different ways of reading are stored as alternative paths,[14] the text of the Semper Edition is represented as a continuous chain of tokens, whose editorial status is indicated by labels and properties, e.g. crossed out words which already have the label "token" are provided with an additional label "del". The text variants can be read from this graph by means of a query, which returns all nodes of a paragraph. The labels and properties can be used to skip certain words. In this process, a variant is not one or more paths provided by the

editors, but a collection of potentially infinite possibilities that can be created by specific filtering.

This representation corresponds to the reality of working with historical source texts, where single interventions can rarely be collated with each other into groups. It also corresponds to the current aspiration of digital scholarly editing to map the fluidity and instability of text.[15] The website of the Semper-Edition represents two possible paths through the text of the sources, corresponding to the categories "transcription" and "reading text" (see below). However, potentially many more variants can be created to answer new research questions from different disciplines. In order to make this possible, the graph database will be archived for the long term, and a current copy will be made available upon request by interested researchers in the near future.

Presentation of the Edition

Traditional classification of editions has been challenged by the possibilities of the digital environment in recent years. The once common differences between diplomatic, semi-diplomatic and reading editions have been negated by the digital medium, since they ultimately indicate merely different ways of representing edited text. The once necessary aim of diplomatic transcriptions to present the page of the source in as much detail as possible has been made obsolete in the context of the web interface, where the facsimile can always be viewed. At the same time, today's digital editions increasingly emphasize flu-
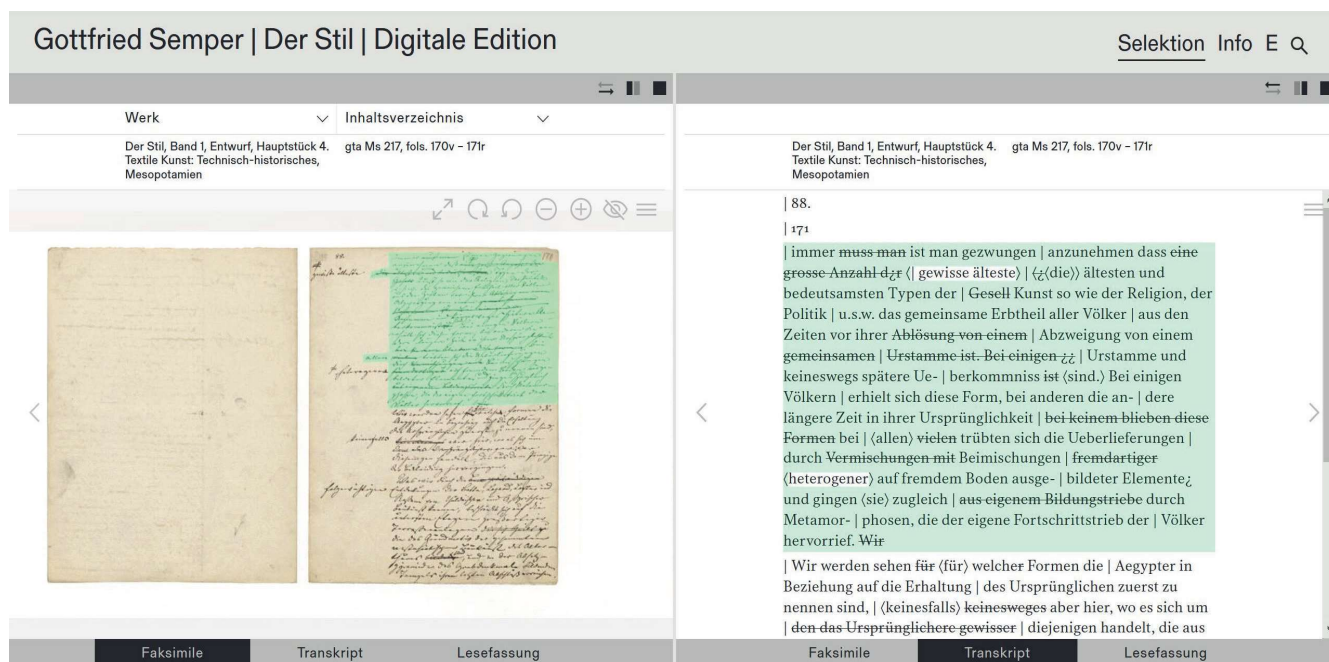

Figure 6.   Text and facsimile shown side-by-side on the webpage of the Semper-Edition.
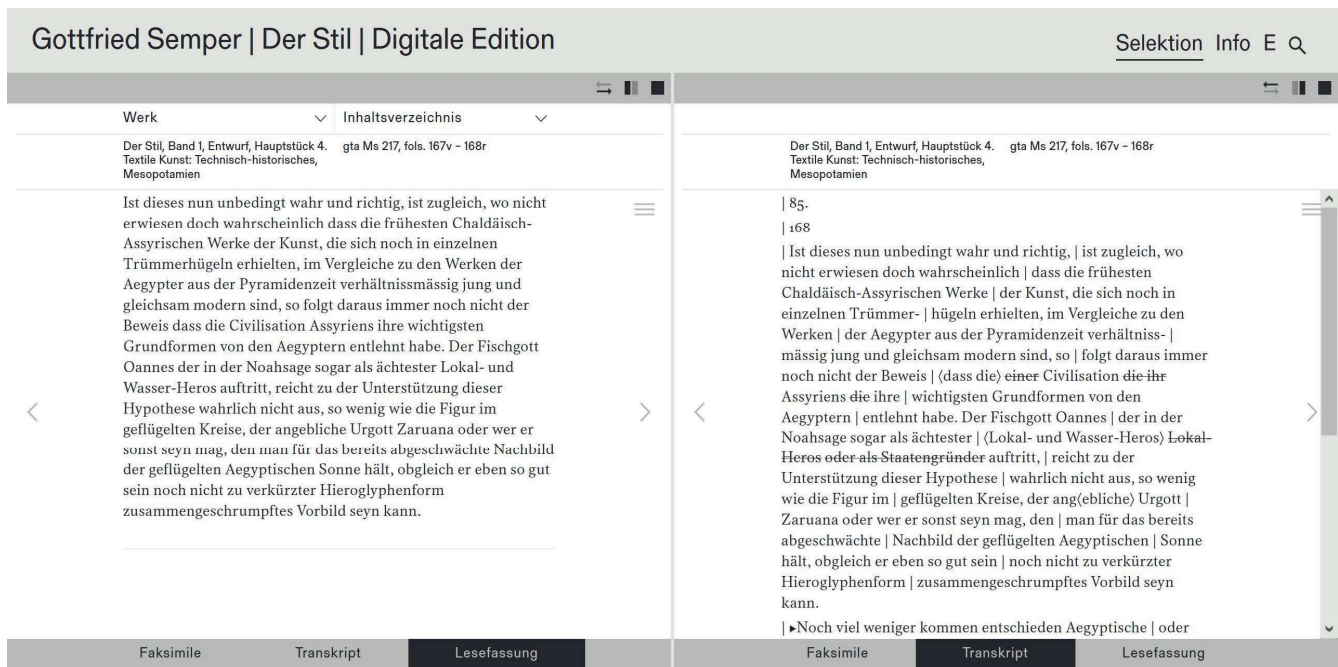
Figure 7 (left panel – Lesefassung):

Werk | Inhaltsverzeichnis

Der Stil, Band 1, Entwurf, Hauptstück 4. Textile Kunst: Technisch-historisches, Mesopotamien — gta Ms 217, fols. 167v – 168r

Ist dieses nun unbedingt wahr und richtig, ist zugleich, wo nicht erwiesen doch wahrscheinlich dass die frühesten Chaldäisch-Assyrischen Werke der Kunst, die sich noch in einzelnen Trümmerhügeln erhielten, im Vergleiche zu den Werken der Aegypter aus der Pyramidenzeit verhältnissmässig jung und gleichsam modern sind, so folgt daraus immer noch nicht der Beweis dass die Civilisation Assyriens ihre wichtigsten Grundformen von den Aegyptern entlehnt habe. Der Fischgott Oannes der in der Noahsage sogar als ächtester Lokal- und Wasser-Heros auftritt, reicht zu der Unterstützung dieser Hypothese wahrlich nicht aus, so wenig wie die Figur im geflügelten Kreise, der angebliche Urgott Zaruana oder wer er sonst seyn mag, den man für das bereits abgeschwächte Nachbild der geflügelten Aegyptischen Sonne hält, obgleich er eben so gut sein noch nicht zu verkürzter Hieroglyphenform zusammengeschrumpftes Vorbild seyn kann.

Faksimile | Transkript | Lesefassung

Figure 7 (right panel – Transkript):

Der Stil, Band 1, Entwurf, Hauptstück 4. Textile Kunst: Technisch-historisches, Mesopotamien — gta Ms 217, fols. 167v – 168r

| 85.
| 168
| Ist dieses nun unbedingt wahr und richtig, | ist zugleich, wo nicht erwiesen doch wahrscheinlich | dass die frühesten Chaldäisch-Assyrischen Werke | der Kunst, die sich noch in einzelnen Trümmer- | hügeln erhielten, im Vergleiche zu den Werken | der Aegypter aus der Pyramidenzeit verhältniss- | mässig jung und gleichsam modern sind, so | folgt daraus immer noch nicht der Beweis ⟨dass die⟩ einer Civilisation die ihr Assyriens die ihre | wichtigsten Grundformen von den Aegyptern | entlehnt habe. Der Fischgott Oannes | der in der Noahsage sogar als ächtester | ⟨Lokal- und Wasser-Heros⟩ Lokal-Heros oder als Staatengründer auftritt, | reicht zu der Unterstützung dieser Hypothese | wahrlich nicht aus, so wenig wie die Figur im | geflügelten Kreise, der ang(ebliche) Urgott | Zaruana oder wer er sonst seyn mag, den | man für das bereits abgeschwächte | Nachbild der geflügelten Aegyptischen | Sonne hält, obgleich er eben so gut sein | noch nicht zu verkürzter Hieroglyphenform | zusammengeschrumpftes Vorbild seyn kann.
| ▶Noch viel weniger kommen entschieden Aegyptische | oder

Faksimile | Transkript | Lesefassung

Figure 7. Two different display modes for the same page of manuscript text.

Figure 8 (left panel – Transkript):

Werk | Inhaltsverzeichnis

Der Stil, Band 1, Entwurf, Hauptstück 4. Textile Kunst: Technisch-historisches, Mesopotamien — gta Ms 217, fols. 167v – 168r

| 85.
| 168
| Ist dieses nun unbedingt wahr und richtig, | ist zugleich, wo nicht erwiesen doch wahrscheinlich | dass die frühesten Chaldäisch-Assyrischen Werke | der Kunst, die sich noch in einzelnen Trümmer- | hügeln erhielten, im Vergleiche zu den Werken | der Aegypter aus der Pyramidenzeit verhältniss- | mässig jung und gleichsam modern sind, so | folgt daraus immer noch nicht der Beweis ⟨dass die⟩ einer Civilisation die ihr Assyriens die ihre | wichtigsten Grundformen von den Aegyptern | entlehnt habe. Der Fischgott Oannes | der in der Noahsage sogar als ächtester | ⟨Lokal- und Wasser-Heros⟩ Lokal-Heros oder als Staatengründer auftritt, | reicht zu der Unterstützung dieser Hypothese | wahrlich nicht aus, so wenig wie die Figur im | geflügelten Kreise, der ang(ebliche) Urgott | Zaruana oder wer er sonst seyn mag, den | man für das bereits abgeschwächte | Nachbild der geflügelten Aegyptischen | Sonne hält, obgleich er eben so gut sein | noch nicht zu verkürzter Hieroglyphenform | zusammengeschrumpftes Vorbild seyn kann.
| ▶Noch viel weniger kommen entschieden Aegyptische | oder

Faksimile | Transkript | Lesefassung

Figure 8 (right panel – Transkript):

Werk | Inhaltsverzeichnis

Der Stil, Band 1, Reinschrift, Hauptstück 4. Textile Kunst: Technisch-historisches, Mesopotamien — gta Ms 233, fols. 93v – 94r

| 94
| Bd. I. s. 270.
| vielleicht erklärt sich die Sache besser | nach der Hypothese der Wanderungen | u des Colonisirens, vielleicht wirkte beides | durcheinander, indess ist man gezwun- | gen anzunehmen, dss gewisse älteste | Typen der Kunst sowie der Religion, | der Politik &sw. das gemeinsame Erbtheil | alter Völker aus den Zeiten vor ihrer Ab- | zweigung von einem Stamme u keines- | wegs spätere Uebereinkomniss sind. Bei | einigen Völkern erhielt sich diese Form | bei andern die andere längere Zeit in | ihrer Ursprünglichkeit, bei allen trübten | sich die Ueberlieferungen durch Beimi- | schungen heterogener auf fremdem Bo- | den ausgebildeter Elemente u gingen | sie zugleich durch Metamorphosen, die | der eigene Fortschrittstrieb der Völker | hervorrief.
| Wir werden sehen für welche Formen | die Aegypter in Beziehung auf die | Erhaltung des Ursprünglichen zuerst | zu nennen sind, keinesfalls aber hier, | wo es sich um diejenigen handelt | die aus dem Prinzipe der Bekleidung | hervorgingen.
| Was wir durch die folgerichtigen Ent- | deckungen der Botta, Layard, Loftus & | Rassam von Chaldäischer & Assyrischer |

Faksimile | Transkript | Lesefassung

Figure 8. Transcriptions of manuscript draft and fair copy of the same section of text.

idity of text. While print editions tended to designate the variant to the notes of the critical apparatus, the possibilities of the digital environment have allowed it to come into its own. Semper-Edition exploits these possibilities by explicitly representing correspondence between the text and the facsimile (fig. 6), by allowing two representations of the edited text – a transcription and a reading version (fig. 7) – and by giving the user an opportunity to compare pairs of freely selected sources in any combination (fig. 8).

The user can choose to display a complete transcript or a reading version. A transcript shows the whole content of the recorded text phenomena. For example, line breaks are marked, deleted parts are shown as such and insertions are

Werk ⌄	Inhaltsverzeichnis ⌄	Index ✕
Der Stil, Band 1, Entwurf, Hauptstück 4. Textile Kunst: Technisch-historisches, Mesopotamien	gta Ms 217, fols. 167v – 168r	**Dokumentinformationen**

| 85.

| 168

| Ist dieses nun unbedingt wahr und richtig, | ist zugleich, wo nicht erwiesen doch wahrscheinlich | dass die frühesten Chaldäisch-Assyrischen Werke | der Kunst, die sich noch in einzelnen Trümmer- | hügeln erhielten, im Vergleiche zu den Werken | der Aegypter aus der Pyramidenzeit verhältniss- | mässig jung und gleichsam modern sind, so | folgt daraus immer noch nicht der Beweis | ⟨dass die⟩ ~~einer~~ Civilisation ~~die ihr~~ Assyriens ~~die ihre~~ | wichtigsten Grundformen von den Aegyptern | entlehnt habe. Der Fischgott Oannes | der in der Noahsage sogar als ächtester | ⟨Lokal- und Wasser-Heros⟩ ~~Lokal-Heros oder als Staatengründer~~ auftritt, | reicht zu der Unterstützung dieser Hypothese | wahrlich nicht aus, so wenig wie die Figur im | geflügelten Kreise, der ang⟨ebliche⟩ Urgott | Zaruana oder wer er sonst seyn mag, den | man für das bereits abgeschwächte | Nachbild der geflügelten Aegyptischen | Sonne hält, obgleich er eben so gut sein | noch nicht zu verkürzter Hieroglyphenform | zusammengeschrumpftes Vorbild seyn kann.

| ▸Noch viel weniger kommen entschieden Aegyptische | oder

Faksimile **Transkript** Lesefassung

Stellenkommentar

Personen (3)
Oannes, mythisches Mischwesen aus Fisch und Mensch
Austen Henry Layard (1817-05-03 Paris - 1894-07-05 London), Archäologe Diplomat
Zaruana Akarana, doppelgeschlechtliche Urgottheit aus der persischen Mythologie

Orte (1)
Mossul (Ninive), Irak (Mesopotamien)

Völker (1)
Ägypter, Aegypter

Zitieren

Metadaten

Figure 9.   Fold-out index with individual comments highlighted with mouse-over.

shown with markers. The reading version is a machine-generated representation of the XML manuscript transcriptions, which shows all deletions and insertions in a continuous text.

The two-part working area of the edition interface in comparison view mode is designed to be as flexible as possible to allow a free choice of documents and ensure maximum comparability of different drafts of *Style*. In the windows on the right and left, any source available as part of the Edition can be selected for display. For more comfortable reading, the single view function is also available. The drop-down menus for contents allow one to select a different source at any time without having to navigate back to the start page. The two menus give the possibility of working with both descriptive headings and archival identifiers. If two views of the same source are selected (e.g. a facsimile and a transcript), they also scroll back and forth synchronously. If pages of a different source document are selected, the scrolling happens separately.

Editorial commentary is displayed in the fold-out "Index" field on the basis of the transcription markup and relevant index entries. The fold-out can be expanded at any time for each page, and also displays the correct citation for the current source. In the future it will contain expanded editorial commentary, metadata and links to equivalent text passages within the Edition, e.g. draft of similar text in a different manuscript. Prospectively, it should also be possible to link to external resources, e.g. artefacts presented on digital museum platforms. The individual comments are linked to their places in the text by highlighting them during mouse-over (fig. 9). In the comparison view, comments are displayed for both

currently opened pages, and their positions are highlighted on both sides to facilitate comparison of the documents.

The Edition webpage at present accommodates a simple search that enables users to find pages containing specific words and phrases within the edition and to use basic functions to combine search terms (e.g. Boolean modifiers).

Future Development

Like any representation of an archive, digital or analogue, Semper-Edition is partial and contingent and must, of necessity, be selective when faced with the vast archives of Gottfried Semper (table 1).

In the future, we plan to expand the content of the Edition, and to enhance its functionality and research potential. In particular, linked data will form one of the main concerns of the Edition in its second four-year phase (pending funding approval), since Semper's texts are extremely rich in references to entities from their textual, conceptual and material contexts, as indicated above. In addition, a complex search function will be made available to enable the user to conduct semantic searches. For example, it would be possible to conduct a search to display all artefacts mentioned within a two-page range of any instance of the term "Bekleidung", or to see which other conceptual terms appear in the same context. At present such a search is offered by very few graph-based front-ends and the usability of querying graph databases on the whole could still be considerably improved. Its development will require a substantial creative input as

Style in Technical and Tectonic Arts, or Practical Aesthetics						
Phase 1 (to be completed December 2020)		Projected Phase 2 (pending funding approval)			Projected Phase 3 (pending funding approval)	
Textile Art I	Textile Art II	Theory of Art Forms (Kunstformenlehre)	Ceramics, Tectonics, Stereotomy, Metalwork I	Ceramics, Tectonics, Stereotomy, Metalwork II	Comparative Architecture (Vergleiche Baulehre)	Lectures
Manuscripts in Semper's hand: Ms 215–Ms 220. Fair copies with Semper's corrections: Ms 229–Ms 235. Print proofs with Semper's corrections: Ms 252. Print variants: 1860 a, b and c; 1878.	Prospectus: Ms 195: Prospectus, fragments of "Prolegomena" and outline. Ms. 196–201: "Prospectus" of vol. 2 of *Style*. Ms 202: Semper's outline of vol. 3 of *Style*. Ms 203: Copy of the outline of vol. 3 of *Style*. Ms 204: Fair copy of the outline of vol. 3 of *Style*. Ms 205: printed prospectus. Ms 184-194: Draft fragments. Variants: Ms 206, Ms 207, Ms 208.	Ms 168–177: Preface and introduction. Ms 178–182: Copy of the version at UB Braunschweig, with corrections for *Style*. Ms 182a: Draft of letter to publisher Vieweg. Ms 183: Fragments. Manuscript from the collection of UB Braunschweig (Vieweg-Archive, V3:1.1.3.32/I, II, III and IV).	Manuscripts in Semper's hand for vol. 2 of *Style*, 1863: Ms 221–Ms 228. Fair copies: Ms 236–Ms 238. Print proofs with Semper's corrections: Ms 253. Print variants: 1863 a, b and c; 1879.	Various drafts: Ms 191: Stereotomy. Ms 192: Ceramics. Ms 209–Ms 210: Ceramics. Ms 211: Tectonics. Ms 212: Stereotomy. Ms 213: Metalwork. Ms 214: Various. Excerpts and Notes: Ms 246–Ms 251.	"Ueber Baustyle" (1869) *Style* vol. 3 (1870–77) Introduction, 1. Ancient Art Comparative Architecture (1849–77). 6. Assyria and Chaldea, 7. Continuation, 8. Ninive, 9. Babylon, 10. Conclusion, 11. Media and Persia, 12. People of East and South Asia, 13. China, 14. Construction and Particularities of China, 15. Hindustan, 16. Indian Architecture, 17. Civil architecture of the Hindus, 18. Construction of the Hindus, 18. Egypt, 19. House building of the Egyptians. Prospectus for a two- and three-volume *Style*. Notes to *Style* vol. 3, excerpts.	"Geschichte der Baukunst" (1861/62), lecture notes by A. F. Bluntschli. "Geschichte der Baukunst" (1862/63), lecture notes by R. Rahn. "Vergleichende Baulehre" (1863), lecture notes by A. Schäfer. "Geschichte der Architektur" (1863–1871), lecture notes by anonymous. "Vergleichende Baukunde" (1863–1871), lecture notes by anonymous. "Styllehre" (1869/70), lecture notes by E. Bumand. "Vergleichende Baulehre" (1870/71), lecture notes by E. Bumand. Lecture notes by Clemens Steiner.

Table 1. Complete and projected content of the Semper-Edition (sources in the gta Archives / ETH Zurich, the ETH Library, the UB Braunschweig and private collections).

well as innovative technical effort, but it would enable the Edition to provide tools for data exploration that would make its potential accessible to non-technical users.

The choice of a graph database to represent a digital edition is a future-oriented decision. This technology can effectively preserve complex data, permit study of textual content in non-predetermined as well as structured ways, and allow exploration not only of the sources themselves

but also of their conceptual context, thanks to the growing possibilities of linked data. On this basis, Semper-Edition looks to provide a resource for future scholars that would go beyond serving as a book-like reference volume and enable exploration of materials in search of new approaches and avenues of research.

AUTOR

Elena Chestnova, PhD, architect and art historian
Institute for History and Theory of Art and Architecture (ISA)
Academy of Architecture Mendrisio, ORCID: https://orcid.org/0000-0002-0654-1313

Swen Vermeul, MSc, environmental and computer scientist
Scientific Information Services
ETH Zurich, Weinbergstrasse 11, 8092 Zürich, swen@ethz.ch, ORCID: https://orcid.org/0000-0002-1307-3052

NOTES

1 See summary in James E. Dobson, *Critical Digital Humanities: The Search for a Methodology*, 1st edition, Urbana 2019.

2 Wolfgang Herrmann, *Gottfried Semper im Exil – Paris, London 1849–1855. Zur Entstehung des "Stil" 1840–1877*, Schriftenreihe des Instituts für Geschichte und Theorie der Architektur an der Eidgenössischen Technischen Hochschule Zürich, Basel 1978. – Wolfgang Herrmann / Gottfried Semper, *Gottfried Semper theoretischer Nachlass an der ETH Zürich Katalog und Kommentare*, Geschichte und Theorie der Architektur, Basel 1981. – Wolfgang Herrmann / Gottfried Semper, *Gottfried Semper: In Search of Architecture* (Consists of material orig, publ. in *Gottfried Semper Im Exil*, Basel 1978 and in *Gottfried Semper: Theoretischer Nachlass an Der ETH Zürich, Katalog Und Kommentare*, Basel 1981), Cambridge 1984. – Susanne Luttmann, "Gottfried Sempers 'Vergleichende Baulehre': eine quellenkritische Rekonstruktion" (Thesis, Eidgenössische Technische Hochschule ETH Zürich, No. 18110, Zurich, ETH, 2008). – Michael Gnehm, Sonja Hildebrand, Dieter Weidmann (eds), *Gottfried Semper: London Writings 1850–1855,* Zürich 2021.

3 Gottfried Semper, print proofs for *Style*, vol. 1, pp. 514–515. gta Archives, ETH Zurich 20-Ms-252-82.

4 Sonja Hildebrand, *Vom klassizistischen 'Abstraktionswesen' zurück 'auf den natürlichen Weg des Sehens' : Gottfried Sempers Hermeneutik antiker Farbigkeit*, in: Martin Döniker, Jutta Müller-Tamm, Friedrich Steinle (eds), Farben der Klassik, pp. 291–312, Göttingen 2016, pp. 304, 308–309.

5 Elena Pierazzo, *Digital Scholarly Editing: Theories, Models and Methods*, Digital Research in the Arts and Humanities, Aldershot 2016, pp. 117–122.

6 James Cummings, *Converting Saint Paul: A New TEI P5 Edition of the Conversion of Saint Paul Using Stand-off Methodology*, in: Literary and Linguistic Computing, vol. 24, no. 3, September 2009, pp. 307–317. – Gabriel Bodard / Juan Garces, *Open Source Critical Editions: A Rationale*, in: Marilyn Deegan (ed.), Farnham, 2009, Text Editing, Print and the Digital World, pp. 83–98.

7 Desmond Schmidt, *Merging Multi-Version Texts: A Generic Solution to the Overlap Problem*, in: Balisage: The Markup Conference, Montréal 2009, accessed April 21, 2020, https://doi.org/10.4242/BalisageVol3.Schmidt01. – Andreas Kuczera, *Digital Editions beyond XML – Graph-Based Digital Editions*, in: Marten Düring et al. (eds), Proceedings of the 3rd HistoInformatics Workshop on Computational History, Krakow 2016, http://ceur-ws.org/Vol-1632/paper_5.pdf. – Iian Neill / Andreas Kuczera, *The Codex – an Atlas of Relations*, in: Zeitschrift für digitale Geisteswissenschaften, May 2019, http://dx.doi.org/10.17175/sb004_008. – Elli Bleeker, Bram Buitendijk, Ronald Haentjens Dekker, *Agree to Disagree: Modelling Co-Existing Scholarly Perspectives on Literary Text*, in: Digital Scholarship in the Humanities, vol. 34, no. 4, December 2019, pp. 844–854, https://doi.org/10.1093/llc/fqz061.

8 Clemens Liedtke, *DH's Next Top Model? Digitale Editionsentwicklung zwischen Best Practice und Innovation am Beispiel des "Corpus Masoreticum"*, in: Christof Schöch (ed.), DHd 2020 Spielräume: Digital Humanities zwischen Modellierung und Interpretation. Konferenzabstracts, Paderborn 2020, pp. 107–111, https://zenodo.org/record/3666690#.XpnIEnLRY2w.

9 Tobias Strauss et al., *CITlab ARGUS for Historical Handwritten Documents*, in: ArXiv:1412.3949 [Cs], December 2014, http://arxiv.org/abs/1412.3949. – Guenter Muehlberger et al., *Transforming Scholarship in the Archives through Handwritten Text Recognition: Transkribus as a Case Study*, in: Journal of Documentation, vol. 75, no. 5, January 2019, pp. 954–976, https://doi.org/10.1108/JD-07-2018-0114. – Tobias Hodel, *Best-practices zur Erkennung alter Drucke und Handschriften. Die Nutzung von Transkribus large und small-scale*, in: Christof Schöch (ed.), DHd 2020

Spielräume: Digital Humanities zwischen Modellierung und Interpretation. Konferenzabstracts, Padeborn 2020, pp. 84–87, https://doi.org/10.5281/zenodo.3666690.

10   Sᴀʀᴀʜ Mᴏᴍʙᴇʀᴛ, *From Books to Collections: Critical Editions of Heterogeneous Documents*, in: Dᴀɴɪᴇʟ Aᴘᴏʟʟᴏɴ, Cʟᴀɪʀᴇ Bᴇ́ʟɪsʟᴇ, Pʜɪʟɪᴘᴘᴇ Rᴇ́ɢɴɪᴇʀ (eds), Digital Critical Editions, Champaign 2014, p. 256.

11   Mᴇᴀᴅʜʙʜ Hᴇᴀʟʏ, *A Comparison of Graph and Collection Based Representations of Early Modern Biographical Archives*, in: Graph Technologies in the Digital Humanities Conference, Vienna 2020.

12   Tᴀʀᴀ L. Aɴᴅʀᴇᴡs / C. Mᴀᴄᴇ, *Beyond the Tree of Texts: Building an Empirical Model of Scribal Variation through Graph Analysis of Texts and Stemmata*, in: Literary and Linguistic Computing, vol. 28, no. 4, December 2013, pp. 504–521, https://doi.org/10.1093/llc/fqt032. – Rᴏɴᴀʟᴅ Hᴀᴇɴᴛᴊᴇɴs Dᴇᴋᴋᴇʀ / Dᴀᴠɪᴅ Bɪʀɴʙᴀᴜᴍ, *It's More than Just Overlap: Text As Graph*, in: Balisage: The Markup Conference, Rockville 2017, https://doi.org/10.4242/BalisageVol19.Dekker01. – Bʟᴇᴇᴋᴇʀ, Bᴜɪᴛᴇɴᴅɪᴊᴋ, Hᴀᴇɴᴛᴊᴇɴs Dᴇᴋᴋᴇʀ, 2019.

13   E.g. Nᴇɪʟʟ / Kᴜᴄᴢᴇʀᴀ, 2019.

14   Bʟᴇᴇᴋᴇʀ, Bᴜɪᴛᴇɴᴅɪᴊᴋ, Hᴀᴇɴᴛᴊᴇɴs Dᴇᴋᴋᴇʀ, 2019.

15   Pɪᴇʀᴀᴢᴢᴏ, *Digital Scholarly Editing*, pp. 25–29; Eʟʟɪ Bʟᴇᴇᴋᴇʀ, Bʀᴀᴍ Bᴜɪᴛᴇɴᴅɪᴊᴋ, Rᴏɴᴀʟᴅ Hᴀᴇɴᴛᴊᴇɴs Dᴇᴋᴋᴇʀ, *From Graveyard to Graph*, in: International Journal of Digital Humanities, vol. 1, no. 2, July 2019, pp. 141–163, https://doi.org/10.1007/s42803-019-00012-w.

LIST OF ILLUSTRATIONS

SUMMARY

This article presents a novel digital edition of Gottfried Semper's magnum opus *Style,* comprising manuscripts and other sources identified as forming main parts of its evolution and contextualizing it within its intellectual milieu both within and outside of Semper's corpus. The Edition's implementation combines the established paradigm of TEI XML with the new modes of representing text made possible by labelled property graph databases. The paper summarizes its workflow and explains its technical processes to enable critical engagement of readers and scholars. It also presents the current and projected content of the Edition and the outlook for its technical and functional development.

RÉSUMÉ

Cet article présente une nouvelle édition digitale du chef-d'œuvre de Gottfried Semper, *Der Stil*. L'édition comprend plusieurs manuscrits et autres sources identifiées comme parties intégrales de l'évolution de l'œuvre et permettant de la contextualiser à la fois dans le corpus sempérien et dans le milieu intellectuel de l'époque. L'implémentation de l'édition combine le paradigme établi TEI XML avec un nouveau mode de représentation du texte rendu possible par une banque de données « labelled property graph ». L'article résume le flux des travaux de l'édition et explique ses processus techniques afin de permettre l'engagement critique des lecteurs et scientifiques. Il décrit non seulement le contenu actuel et envisagé de l'édition mais aussi les perspectives de son développement technique et fonctionnel.

ZUSAMMENFASSUNG

Dieser Artikel stellt eine neuartige digitale Edition von Gottfried Sempers Hauptwerk *Der Stil* vor. Ediert werden Manuskripte und weitere Quellen, die die Werkgenese in ihren wesentlichen Bestandteilen nachvollziehbar machen. Ausserdem erlauben sie eine Verortung des *Stils* in Sempers gesamtem schriftlichem Werk und in seinem intellektuellen Umfeld. Die Implementierung der Edition kombiniert das etablierte TEI-XML-Paradigma mit einem neuen Modus der Textdarstellung, der durch die Anwendung einer Labelled-Property-Graph-Datenbank ermöglicht wird. Das Paper resümiert den Workflow der Edition und erklärt ihre technischen Prozesse, um Lesern und Wissenschaftlerinnen eine kritische Auseinandersetzung damit zu ermöglichen. Es beschreibt den aktuellen und zu erarbeitenden Inhalt der Edition sowie ihre künftige technische und funktionale Entwicklung.

RIASSUNTO

Questo saggio presenta una nuova edizione digitale dell'opera principale di Gottfried Semper, *Der Stil.* L'edizione include manoscritti e altre fonti, identificati come componenti essenziali per comprendere la genesi dell'opera. I documenti elencati permettono inoltre di collocare *Der Sti*l all'interno dell'opera scritta di Semper e nel contesto intellettuale del suo tempo. L'implementazione dell'edizione combina il paradigma consolidato di TEI XML con le nuove modalità di rappresentazione del testo, rese possibili dall'utilizzo di una banca dati «labelled property graph». Il saggio riassume il lavoro svolto per l'elaborazione dell'edizione e ne spiega i processi tecnici per consentire a lettori e studiosi un'analisi critica. Inoltre, l'articolo descrive i contenuti attuali e di prossima pubblicazione dell'edizione e il suo potenziale sviluppo tecnico e funzionale..