

# BIBLindex

## International Workshop on Computer Aided Processing of Intertextuality in Ancient Languages

2<sup>nd</sup>-4<sup>th</sup> June 2014

Bâtiment Blaise PASCAL, INSA,  
Campus de la Doua, Villeurbanne

Tram: T1/T4, station "Doua Gaston Berger"

Room 501.337, on the 3<sup>rd</sup> floor



gesteert von  
Bundesministerium  
für Bildung  
und Forschung

ISTORRE  
SOURCES  
SOURCES  
SOURCES

LIRIS

GCDH GÖTTINGEN CENTRE FOR  
DIGITAL HUMANITIES

Program: <http://www.bibindex.hypotheses.org/1686>

Contact: [laurence.mellerin@mom.fr](mailto:laurence.mellerin@mom.fr)



Library of Congress, Rosenwald 4, Bl. 5r



## **International Workshop on Computer Aided Processing of Intertextuality in Ancient Languages, 2nd-4th June 2014,**

*Bâtiment Blaise PASCAL, INSA,  
Campus de la Doua, Villeurbanne (France)*

*coorganized by HiSoMA (UMR 5189, Lyon), LIRIS (UMR 5205, Villeurbanne) and by the  
Göttingen Centre for Digital Humanities (e-TRAP), with the support of the National Research  
Agency (ANR Biblindex) and the Partner University Fund (PUF).*

This workshop was initiated as the conclusive meeting of the project Biblindex, funded by the French National Research Agency (ANR), which aims at establishing an exhaustive statement of the biblical references found in the texts of the Late Antiquity and the Middle Ages. At this meeting will be gathered computer scientists and digital humanists, specialists of corpora written in ancient languages. The planned sessions aim to present the state of art regarding concepts and technics used to process quotations in ancient languages. A lot of projects work nowadays on various corpora, asking similar questions about text-reuse. Comparing experiments, we hope to clear perspectives to mutualize developments and methodological choices, in order to build a federative project at the European scale in the coming years.

The first session will be devoted to mutual project presentations. Afterwards, the various stages of quotation processing will be discussed in four workshops. The first two of them will tackle the preparation of sources and the automatic retrieval of concordancing text places: interests and complementarities of statistical and linguistic approaches will be compared. The next two will focus on the conceptual definitions, the modelling of the unstable idea of “quotation” and the XML-TEI encoding to implement for its characterization, in close interdependence with visualization choices.

### **ORGANIZERS:**

**Marco BÜCHLER** (GCDH), **Elöd EGYED-ZSIGMOND** (LIRIS), **Laurence MELLERIN**  
(HiSoMA)



## PROGRAM

---



---

**Monday, 2<sup>nd</sup> June**

### **SESSION 1: PROJECTS OVERVIEW**

This morning will be widely open to the public, both humanists and computer scientists. This session aims to inform the participants about some bigger projects being available. Presentations of the projects give an overview to the research questions, used data, and project objectives.

**8:45** Welcome

**Moderator: Yasmine ECH CHAEL (HiSoMA, Lyon)**

### **BIBLICAL QUOTATIONS IN PATRISTIC TEXTS**

**9:00** **The Biblindex Project, Index of Biblical Re-uses in the Early Christian Literature (Laurence MELLERIN, HiSoMA, Lyon)**

The BIBLINDEX project, funded by the French National Research Agency and supported by the Institut des Sources Chrétiennes in Lyon, helped initiate the creation of an index of biblical quotations and allusions in Early Christian Literature, which plans to become exhaustive. It seeks to link a corpus of biblical texts – collections of scriptural books which were originally written in various languages and translated early in their history – with a corpus of ancient and medieval authors, who refer to the Bible as a fixed entity yet at the same time contribute through their quotations to the form and concept of ‘the text’.

The first step of our enterprise was to establish a nomenclature for our sources, e.g. providing keys to precise lists of authors and works, instituting fixed points of reference and the identification of reliable critical editions and other types of relevant material. Then biblical text re-uses have been

gathered. At present, BIBINDEX consists of a comprehensive inventory of 700.000 biblical quotations and allusions in Early Greek and Latin Christian Literature. The database now includes eleven biblical texts written in different ancient and modern languages. A multilingual concordance between those Bibles has been created, allowing the user to visualize the biblical text in order to compare it with the quotations found. We already propose simple search forms to access to the data on the website <http://www.bibindex.org>.

Each entry on the website offers a series of numbers indicating the chapter and verse of the biblical text, its location in the patristic writing and the corresponding page and line numbers in the reference edition. We now need to make links between the BIBINDEX database and the existing textual databases.

**9:30 The Digital Greek Patristic Catena (Athanasios PAPARNAKIS, Aristotle University, Thessaloniki)**

Digital Greek Patristic Catena: Based on a collection of nearly 350.000 biblical quotations in the Patristic texts from Migne's *Patrologia Graeca* edition of nearly 900 ecclesiastical authors and 6000 works, a database has been developed following the methodology of the ancient hermeneutical tool of the catenae. The database utilises four sets of information:

- a) the Greek text of the Bible and modern translations,
- b) the authors and works,
- c) the patristic text in two forms: short patristic passages containing the biblical quotations and an image of the full page and
- d) indexes of subjects, names and contents.

**9:45 The Vetus Latina Iohannes and COMPAUL Projects (Catherine SMITH; Rosalind MACLACHLAN, ITSEE, Birmingham)**

Early Christian Latin writers frequently cite, adapt and discuss the text of the Bible; since they are using early forms of the biblical text, their evidence can provide important information about that text as well as offering insight into how the Bible was read and interpreted in the early period of its history.

The *Vetus Latina Iohannes* Project is producing an edition of the Old Latin materials for the Gospel of John that provide evidence for early alternatives and predecessors to today's familiar Vulgate Latin translation; two fascicles have been published so far covering John 1-9. There are two sorts of materials: manuscripts with elements of pre-Vulgate textual traditions and citations from patristic authors. The project thus has two technical components: an electronic edition of the manuscripts and a database of citations of the Gospel

of John in patristic works. Data from both of these is brought together in the final printed edition.

The project has made electronic transcriptions of the Gospel manuscripts containing texts with a potential Old Latin element following the TEI guidelines with specific encoding of features found in biblical manuscripts. The online edition of these manuscripts (<http://www.iohannes.com/vetuslatina>) allows access to both transcriptions of individual manuscripts and a synopsis of each verse in all the manuscripts. In the case of the Gospel of John in the Old Latin Bible, the main strands of Old Latin text are represented by extant manuscripts and the transcriptions of these manuscripts thus provide a framework to add the patristic citations.

For the past century details of patristic citations and references have been manually collected from printed editions on index cards by the Vetus Latina Institute at Beuron; these Beuron cards, some typed, some handwritten, have now been digitally imaged. The *Vetus Latina Iohannes* Project has transcribed around 60,000 citations for the Gospel of John from these cards and other sources into Excel spreadsheets which have now been transformed into an online database with related details of authors, works and editions. Corraling the diverse and sometimes awkward material into a useable database, while maintaining compatibility with legacy systems of referencing, is not without its challenges. From the material in the database, the edition presents citations of each verse and an index of readings for each verse.

The COMPAUL project is investigating the earliest commentaries on the Pauline Epistles as sources for the biblical text. There are an exceptional number of surviving commentaries on the Pauline Letters, in Greek and especially in Latin, composed in the 4<sup>th</sup> Century and including works by significant figures of the period such as Augustine, Jerome and Pelagius. These provide significant evidence for pre-Vulgate textual traditions earlier than those represented in surviving manuscripts. This project thus provides a valuable preliminary to future editions of the Pauline Epistles.

The project has been analyzing the biblical text in each commentary using electronic transcriptions of commentary manuscripts and biblical manuscripts to identify and explore each author's text of the epistles. This involves in particular comparing the biblical text found in the lemma of the commentaries with the references to the biblical text in the authorial exegesis since later generations may have replaced or adapted the biblical text, especially in the lemma, to forms current in their time. This data is being collected in an extension to the patristic database developed for the *Vetus Latina Iohannes* project. We are also starting to explore whether

characteristic readings in the biblical text can be used to trace later authors using a commentary rather than a biblical manuscript.

**10:15 The Digital Processing of Patristic Citations for the Editio Critica Maior [ECM] of Acts (Volker KRÜGER, Gunnar BÜSCH, INTF, Münster)**

The ECM-project aims for a new reconstruction of the initial text of the Greek New Testament and the critical edition of the Acts of Apostles as part of this project is currently nearing completion. This project overview will cover a searchable online database for patristic citations established explicitly for the work on this edition as well as a user interface that allows relating the gathered patristic data with the manuscript data prepared for the critical apparatus.

**10:45 Break**

---



---

**TEXT RE-USE IN THE DIGITAL HUMANITIES**

**11:00 eTRACES, eTRAP (Marco BÜCHLER, GCDH, Göttingen)**

The presentation gives an overview of the recently planned activities of the eTRACES project and the eTRAP Early Career Research Group. The starting point is a requirements analysis for the mining and retrieval tool. Furthermore, the presentation introduces the TRACER framework. This is followed by a brief overview of the use case scenarios.

**11:30 The Leipzig Open Fragmentary Texts Series [LOFTS] (Monica BERTI and Greta FRANZINI, University of Leipzig)**

The goal of this paper is to present a project of the Humboldt Chair of Digital Humanities at the University of Leipzig: the *Leipzig Open Fragmentary Texts Series* (LOFTS). LOFTS establishes open editions of ancient works that survive only through quotations and text re-uses in later texts (i.e., those pieces of information that humanists call “fragments”). LOFTS has two goals: 1) digitize paper editions of fragmentary works and link them to source texts; 2) produce born-digital editions of fragmentary works. LOFTS uses both XML and RDF, the CTS/CITE Architecture, and different data models like the PROV-O ontology, the Systematic Assertion Model (SAM), and the Open Annotation (OA) Core data model.

LOFTS has three submain projects: 1) the *Digital Fragmenta Historicorum Graecorum* (DFHG); 2) the *Digital Marmor Parium* (DMP); 3) the *Digital Athenaeus*.



**12:00 The Sharing Ancient Wisdoms [SAWS] Project: Aims, Problems and Achievements (Charlotte ROUECHÉ, King's College, London)**

The Sharing Ancient Wisdoms project, funded by HERA, ran from 2010 to 2013. Our aim was to explore ways in which to present and analyse citations; we focused on the later Greek tradition, and translations into Arabic. We published texts in TEI compliant XML; we developed CTS identifiers; and we started building an ontology to express the various kinds of relationships with which we were dealing. Above all, we aimed to connect citations – both in collections, such as the *gnomologia* edited by Denis Searby and the Swedish team) or embedded in continuous texts (Kekaumenos, edited by Charlotte Roueché) with one another, and with their source texts. Elvira Wakelnig, and the Vienna team, further explored the relationships of Greek to Arabic collections and Wisdom Literature, and the further journeys of such materials into Latin and Spanish.

This was only a first attempt; but we have made all our materials, and most importantly, our ontology, available, for others to use. This paper will present our results; we hope to discover other colleagues who can make use of our work, and take it forward.

**12:30 Lunch**

**14:00 Corpus der Arabischen und Syrischen Gnomologien [CASG] (Norman WETZIG, Halle)**

The project “*Corpus der arabischen und syrischen Gnomologien*” (CASG) was funded by the Fritz Thyssen Stiftung (2010-2012) and is supervised by Dr. Ute Pietruschka, who together with her assistants will set up a database and digital repository containing all available collections of *Gnomologia* written in Arabic and Syriac.

In the field of literature, Late Antiquity marks a shift in literary style with a preference for encyclopedic works, consisting of summaries of earlier works. Collections of sayings like the *Gnomologium Byzantinum*, the *Gnomologium Parisinum* and the Christian collection of John of Damascus were a popular kind of literature and represented an important source for the transmission of knowledge. Their popularity can be seen as connected to the compilation of anthologies and handbooks on varying topics that were intended for a quick orientation in a certain field of knowledge. A *gnomologium* is a collection of short sentences and anecdotes attributed to more or less commonly known philosophers, poets, rulers, politicians. These collections sometimes are arranged by alphabetical order of the authors, some are arranged by topics while others seem to have no internal

order at all. Most gnomologia are no pure gnomologia as such: amongst real gnomai (sentences), apophthegms, chreiai or diatribes can be found. The gnomological tradition was not limited to Byzantine literature – on the contrary it can be found in all Mediterranean civilizations and those under their cultural influence. We thus find Greek gnomological material even in Coptic and Ethiopic manuscripts. The Arabic gnomologia are the largest extant group of collections beside the Greek ones. The project aims at illuminating the transmission of Syriac and Arabic gnomologia, which have been studied even less than the Greek collections and at comparing them regarding their different aspects.

---

---

## SESSION 2: NATURAL LANGUAGE PROCESSING METHODS FOR RETRIEVING TEXTS AND COMPUTING TEXT SIMILARITIES

**Moderator: Marco BÜCHLER (GCDH, Göttingen)**

**14:30 Introduction: Overview of the Possible Methodologies used in  
Text Retrieval (Marco BÜCHLER)**

With the emerging amount of available digital data, there is a need for more automatic methods. This session introduces work that is related to NLP methods that are relevant in the case of Big Data. In this respect, the computational complexities are highlighted and discussed. The presentation ends with a list and explanation of some NLP problems connected with Big Data.

**14:45 Was it Better Before? Automated Quotation Detection in  
Ancient Texts (Samuel GESCHE and Előd EGYED-ZSIGMOND, LIRIS)**

This work focuses on refining and finding quotations of Greek Church Fathers within the Greek versions of the Tanakh and the New Testament. Our corpus features over 700 works of various Church Fathers, which amounts to around ten millions words.

In order to reach our aim, we had to explore the notion of quotation as relevant to the first centuries of our era, and to discuss the efficiency and usability of modern digital approaches, including their evaluation metrics. We mainly studied unsupervised approaches, either statistical, statistico-structural and statistico-semantic. We ended up building a generic and flexible solution based on a robust document model and an algorithm merging several other approaches. We are currently working on implementing this solution within the GATE framework to facilitate reusability.

In this presentation, we will talk about aspects of the challenges, the model we built to ensure genericity and flexibility, the algorithm we built from the various approaches we explored and discuss our results applying

this algorithm to two different control sets.

**15:15 QuotationFinder – Searching for Quotations and Allusions in Greek and Latin Texts (Luc HERREN, Münster – Londres)**

Traditional search functions of software used to access the TLG (Thesaurus Linguae Graecae) or the CLCLT (CETEDOC Library of Christian Latin Texts) are not well suited for finding quotations and allusions. A search with a Boolean "or" would yield too many matches, as most texts contain at least one of the common words in our search string. A search with a Boolean "and," however, would yield too few, as ancient authors often left out words when quoting without us having a chance to know in advance which ones they would keep, resulting in matches being missed because they lack just one (possibly dispensable) word in our search text.

QuotationFinder gets around this problem by using more sophisticated criteria for determining if a given text is a quotation/allusion. It reads text files exported from the TLG or the CLCLT and considers 5 parameters when it encounters words from your search text: It looks at the number of words matched within a reasonable number of lines; whether the exact form of a word is matched, or a different form of the same word, or a form of a cognate; how close the matched words are to each other; how rare the matched words are; and, finally, to what degree the sequence of the words matches your search text. QuotationFinder then produces an ordered list with the most exact quotation first and the loosest verbal parallel last.

**15:45 Modeling the Scholars: Detecting Intertextuality through Enhanced Word-Level N-Gram Matching in The Tesserae Project, Intertextual Analysis of Latin Poetry (Neil COFFEE, University Buffalo)**

The study of intertextuality, or how authors make artistic use of other texts in their works, has a long tradition, and has in recent years benefited from a variety of applications of digital methods. This paper describes an approach to detecting the sorts of intertexts that literary scholars have found most meaningful, as embodied in the free Tesserae website <http://tesserae.caset.buffalo.edu/>. Tests of Tesserae Versions 1 and 2 showed that word-level n-gram matching could recall a majority of parallels identified by scholarly commentators in a benchmark set. But these versions lacked precision, so that the meaningful parallels could only be found among long lists of those that were not meaningful. The Version 3 search described here adds a second stage scoring system that sorts found parallels by a formula accounting for word frequency and phrase density. Testing against a benchmark set of

intertexts in Latin epic poetry shows that the scoring system overall succeeds in ranking parallels of greater significance more highly, allowing site users to find meaningful parallels more quickly. Users can also choose to adjust recall and precision by focusing only on results above given score levels. As a theoretical matter, these tests establish that lemma identity, word frequency, and phrase density are important constituents of what make a phrase parallel a meaningful intertext.

**16:15 Break**

**16:30 Round-table Discussion: Methodological Discussion about Statistical Approaches**

**Leader: Marco BÜCHLER**

Processing natural language means dealing with the complexity of human interaction. This includes language evolution, the many different dialects, but also any kind of error or change in purpose of meaning. This complexity brings a need for normalization. What is a good level of normalization? If the text is kept to an under-normalized level, then relevant text re-use remains undiscovered. If the text is over-normalized, it becomes easy to spot irrelevant noise. Together with this topic of discussion, the round-table shares experiences and lessons learned.

**18:00 End**



---

---

**Tuesday, 3<sup>rd</sup> June**

**SESSION 3: LINGUISTIC APPROACHES / LANGUAGE SPECIFICITIES**

**Moderator: Guillaume Bady (HiSoMA, Lyon)**

**9:00 Introduction (Pierre-Édouard Portier, LIRIS)**

**9:15 SHEBANQ and Related Projects: Exploring New Directions in the Computational Analysis of Syriac Texts (Wido van Peursen, VU Amsterdam, Eep Talstra Centre for Bible and Computer (ETCBC))**

Since the 1970s, the Eep Talstra Centre for Bible and Computer (ETCBC; formerly known as the *Werkgroep Informatica Vrije Universiteit*) has invested in building a database of the Hebrew Bible incorporating linguistic information at the level of words, phrases, clauses and clause relations. Since 1999, when the ETCBC joined forces with the Peshitta Institute Leiden, the analysis of Syriac texts came into focus and two successful research projects funded by the Netherlands Organization for Scientific Research dealt with multilingual comparison of Hebrew, Syriac and Jewish Aramaic Bible texts: CALAP (1999-2005) and Turgama (2005-2010).

Whereas the ETCBC started as the work of lonely pioneers, the emergence of Digital Humanities research centers and the high position of DH on research agenda's opens up a new potential for the computational exploration of Hebrew and Aramaic text resources, combining various methods of text comparison and text clustering based on linguistic features (comparing syntax trees; vocabulary analysis; statistical glossing) and uninformed or 'blind' techniques (*n*-gramming; Normalized Compression Distance; Kolmogorov complexity). We are now running a pilot to test these methods by training the computer to identify OT quotations in the NT. This might be a first step towards a contribution to a project that aims at recognizing biblical quotations in larger corpora.

**9:45 GREgORI: Softwares, Linguistic Data and Corpus for Ancient GREEk and ORiental Languages (Tamara PATARIDZE; Emmanuel VAN ELVERDINGHE, Université catholique de Louvain, Institut orientaliste)**

The *Research Project in Greek Lexicology* (UCL: Prof. B. Coulie, Dr. B. Kindt) pursues a twin goal: on the one hand, constituting an electronic dictionary of Ancient and Byzantine Greek, on the other hand, producing lemmatized concordances. The dictionary gathers linguistic data directly stemming from corpus-based observations, without any restriction regarding the handled texts' date, literary genre, language level or dialect; moreover, every occurring word is recorded. Each corpus analysis thus provides a comprehensive lexical inventory of the processed texts, lemmatized and perfectly disambiguated. On this base are produced various lexicological tools: lemmatized concordances, lemmatized indexes, reverse indexes, frequency indexes, index of words common or specific to two corpuses or corpus parts.

The know-how developed for Greek is now being extended to the other languages of Christian Orient within the framework of the *GREgORI: Softwares, linguistic data and tagged corpus for Ancient GREEk and ORiental languages* project, the aim being as well unilingual as multilingual. Emmanuel Van Elverdinghe studies the formulaic style found in Armenian colophons: the purpose is to navigate through a corpus spanning several centuries in order to extract stereotypical lexical and syntactic patterns, whose repeating or varying provide information concerning textual transmission as much as the copyists' customs. Tamara Pataridze works out bilingual Greek-Georgian lemmatized lexicons of Gregory Nazianzen's *Discourses*: a processing inspired by the alignment methods leads to accurate analysis of the translation techniques used by the authors of the Oriental versions of the Theologian's *Discourses*.

**10:15 Utilisation de la plateforme d'élévation de données DataLift dans un cadre linguistique, application à l'arménien classique (Gabriel KEPEKLIAN, ATOS Origin)**

La plateforme DataLift (<http://www.datalift.org/>) est le résultat d'un projet ANR-Contint qui vient de se terminer. Il s'agit d'une solution « tout en un » : bâtie sur une socle technique conçu pour une forte modularité. La métaphore de l'ascenseur (Lift) correspond au processus d'élévation qui se décompose en 5 étages. Le premier est destiné à la capture de jeux de données structurés mais hétérogènes et à la désignation d'ontologies permettant



de les décrire. Le deuxième étage comprend les convertisseurs/mappeurs qui appliquent les ontologies et produisent des jeux de données RDF. Au troisième étage, ces jeux élevés (ils ne sont plus englués dans la gangue de leurs formats d'origine) sont stockés dans un triple store. Le quatrième étage est dédié à l'interconnexion de données : en mettant à profit les alignements que les ontologies mobilisées au 2<sup>e</sup> étage, on peut produire de nouvelles données. Le dernier étage est dévolue à l'exploitation des données élevées et interconnectées. Si l'on se rapporte à l'échelle proposée par Tim berners-Lee (<http://www.w3.org/DesignIssues/LinkedData.html>, 2), relative aux données du « web of data », DataLift permet de passer aux données 5 étoiles.

L'exploitation des données peut se faire à l'aide de requêtes exprimées en SPARQL, c'est-à-dire dans une modalité logique très expressive. Il est par exemple possible de faire de l'inférence. Une telle plateforme peut donc être utilisée dans le cadre de la linguistique si l'on ramène la problématique à des jeux de données structurées et des requêtes logiques (logique de 1<sup>er</sup> ordre).

C'est ce que nous avons fait pour une première approche du traitement de l'arménien ancien. Le développement était très léger, pas plus de quelques heures et le résultat est très encourageant (<http://www.kepeklian.com/blog/2014/02/27/analyse-grammaticale-armenien-classique-datalift/>, 3). La même approche peut être reproduite à l'envi pour d'autres langues. Nous avons créé un tokenniseur unicode de l'arménien classique, constitué une base explicite de lemmisation et utilisé un dictionnaire de traduction arménien / français. Nous avons donc trois jeux de données sous forme de fichier : premièrement les données produites par le tokenniseur ([http://www.kepeklian.com/tokenisation\\_arm/tokenisation2.php](http://www.kepeklian.com/tokenisation_arm/tokenisation2.php), 4) à partir d'un texte brut, deuxièmement la base de lemmisation, troisièmement le dictionnaire. Les ontologies sous-jacentes sont produites « à la volée » par la plateforme lors de la capture des jeux de données. Nous montrerons quelques requêtes SPARQL qui permettent de répondre à des questions d'analyse grammaticale du texte, ou de traduction.

**10:45 Break**

**11:00 The Text Alignment Protocol, a Proposed Data Model for Interchangeable Bitexts (Joel KALVESMAKI, Dumbarton Oaks)**

The Text Alignment Protocol (TAP for short) is a nascent XML data model and set of recommended best practices to serve many people who wish to encode, exchange, and study texts and their translations, quotations, paraphrases, adaptations, and summaries. TAP files are designed to be maximally readable and editable by both humans and

machines, independent of (but compatible with) third-party software. The language-agnostic protocol is modular, allowing it to extend and grow, and permitting editors and researchers to work independently, collaboratively, and within their preferred assumptions and purposes. Although expressive of scholarly nuance and complexity, TAP files are meant to benefit scholars and nonscholars alike—anyone interested in the detailed study of ancient, medieval, and modern translations, paraphrases, and related textual reuse.

In this presentation I introduce the TAP data model by discussing its structure, the envisioned distributed workflow, and the four modules that have so far been developed: transcriptions (modified TEI), alignment, lexicomorphology, and grammar. I focus on the principles that have guided TAP design, and discuss challenges and successes that beset a search for a way to encode texts that synthesizes the simplicity cherished by computer scientists, the complexity treasured by philologists, and the self-awareness prized by theorists.

I summarize creative applications that TAP could serve, i.e., multilingual publishing, language learning, and semi-automated machine translation. But I also conclude with a summary of challenges that remain, not least of which is the lack of a broad range of collaborators.

### **11:30 Round-table Discussion: the Relationship between the Concept of Language and Linguistic Approaches**

**Leader: Sara SCHULTHESS (University of Lausanne)**

- Lexical, semantic approaches: lemmatization, standardization: to what extent is each language specific?
- Syntactic approaches: to what extent is each syntax specific?
- Articulation of different linguistic approaches
- What can be reused from one language to another? Step by step, which tools could be thought generic?

**12:30 Lunch**

---

---

## **SESSION 4: TOWARDS A DIGITAL ECOSYSTEM FOR TEXT RE-USE AND ITS APPLICATIONS**

**Moderator: Elöd EGYED-ZSIGMOND (LIRIS, Lyon)**

**14:00 Introduction - Requirements for a Digital Ecosystem (Marco BÜCHLER, GCDH)**

A lot of text re-use has been discovered over the last centuries being re-published in books. Some of it kept existent while others got lost. In a digital world, data can be easily copied so that the information loss can be more or less ignored. What are the requirements for a digital system as this? This brief introduction gives an overview to some requirements on different levels.

**14:15 TEI Encoding Patterns for Citations, Quotations and Allusions: What's "in Stock" ? (Emmanuelle MORLOCK , HiSoMA)**

The concept of citation implies a decontextualization and an incorporation, with different degrees of accuracy and explicitness. In a practical perspective, representing this incorporation with markup means aligning or combining the logics of two texts. It also requires to take pay special attention to the explicitness of the encoding and the disambiguation of implicit information, to permit for example comparative extractions for intertextuality studies. A great deal of flexibility and a global view of what's available in the encoding scheme is thus needed. The talk will give a synthetic view of the main mechanisms offered by TEI guidelines. The examples used will be taken from encoding samples realized in the context of the Biblindex project.

**14:30 Using CTS for Hierarchical Annotations (Chris BLACKWELL, Neel SMITH, Holy Cross College and Furman University) Skype session**

Originally designed to meet the needs of the Homer Multitext project, the CITE architecture defines URN notations for two fundamental kinds of scholarly citation: the CTS URN for textual citation and CITE Collection

URN for citing other kinds of discrete objects. Objects in CITE Collections may or may not be ordered; in contrast, citable nodes of text are always ordered and situated in both a citation hierarchy and a work hierarchy.

Software for working with data in the CITE architecture includes a suite of network services for retrieving objects identified by URN, and utilities for automatically building an RDF graph of all relations specified by URN citation from an archive of CTS and CITE Collection data in simple text formats. (See <http://cite-architecture.github.io/>)

In this paper, we show how URN-aware applications can use such a scholarly graph to align any kind of analytical data set with textual citation so that analytical data can be identified and retrieved in terms of the order and hierarchy of textual citation. We will illustrate how citation of analytical data as a further hierarchical extension of textual citation can capture many kinds of intertextual relations, including: “fragments” and quotations of texts of one text by another; direct commentary by one text on another; shared physical and paleographic features of different texts; relations of features of diagrams to accompanying texts; and comparable geographic, temporal and quantitative contents of different texts.

### **15:00 Declaring Quotations through Canonical Reference Numbers in the Text Alignment Protocol (Joel KALVESMAKI, Dumbarton Oaks)**

In this presentation, I show how the Text Alignment Protocol (TAP) may be used to encode quotations (verbatim or not) that one work makes of another. I use as a simple example Evagrius Ponticus’s *Practicus*, a fourth-century Greek text about monastic asceticism. The short preface, my focus here, quotes five times from the Scriptures, including once from the Psalter, and so illustrates the challenge of how to encode a quotation from a work that has alternative canonical numbering schemes.

I survey how the challenge can be tackled natively in standard Text Encoding Initiative (TEI) files. I argue that key problems beset the standard TEI model: ambiguity, siloed (single-project) terminology, and lack of nuance. I then show how the issue is handled in TAP files: the quoted work (the target) and the quoting work (the source) are transcribed in separate TEI files, customized to be restricted to a single version of a work, structured according to a single canonical reference system that has a typology. Between the source and target stands a simple alignment file that declares the distance between the two, the types of textual reuse (in this case quotation), and the corresponding canonical references between the target and the source. I illustrate ways that doubt, ambiguity, and changing text reuse strategies can be declared in a TAP alignment file. Using the example of the Psalter, I

also show how TAP can be used to reconcile conflicting canonical reference systems for the same work.

**15:30 Visual Quotation (Michèle BRUNET, HiSoMA)**

The workshop will be devoted to the exploration of the concept of “quotation” and to all the digital processing arising from the definitions. Intertextuality, reuse, different degrees of duplication (literal quote, explicit quote, more or less involuntary quote), different degrees of accuracy, of imitation, or pastiche : all the materials which support this survey will be texts, and for the most, ancient texts.

It could be interesting to show that quite the same mechanisms are at work in the use of images and that the typology we try to build for text-materials may be transposed to images. Furthermore, sometimes, both textual and visual citation are used together, the two ways of quoting reinforcing each other.

How can we use digital tools — and which tools ? — to search for and to analyze visual quotation ?

Some examples taken from the Collection of Greek inscriptions of the Louvre Museum will support the analysis.

**16:00 Break**

**16:15 Methodology for LOFTS Fragments: the DFHG Project and the Digital Marmor Parium Project (Monica BERTI, University of Leipzig)**

DFHG is producing a digital edition of the five volumes of Karl Müller’s *Fragmenta Historicorum Graecorum* (FHG) (1841-1870) according to the EpiDoc Guidelines and the CTS/CITE Architecture. This project has produced a catalog of more than 600 fragmentary authors edited by Müller, which is part of the Perseus Catalog, and a set of Guidelines that are contributing to the Epidoc community.

DMP is working on a digital edition of the so called *Marmor Parium*, a Greek universal chronicle on a marble stone.

**16:45 Round-table Discussion: Is a Common Methodology Possible?**

**Leader: Greta FRANZINI (Univ. Leipzig)**

There are several ways in which to store text re-use data, none of which are considered the standard. However, at this stage, in order to come up with one global Digital Ecosystem, it has become necessary to set a common standard.

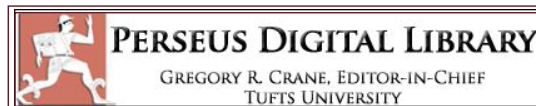
The round-table will, for example, discuss the pros and cons of in-document vs. stand-off markup. In addition, there will be questions to address, such as, what would the digital ecosystem look like: one local repository, or better a distributed infrastructure? What are good common standards for the communication between existing and future projects?

**17:45** End

**19:45** Public lecture : June 1914 and Philology for the  
21st Century (**Gregory CRANE, Perseus Project, Tufts/Leipzig**)

**MOM, Amphithéâtre Benvéniste,  
5/7 rue Raulin, 69007 Lyon**

We are approaching the one hundred year anniversary of the First World War. Philology was never more prestigious than it was a century ago but what, if anything, did it do to counteract the threads of nationalism that led to the disasters of the twentieth century? Far as Europe has come a hundred years later, the vast majority of those who engage with historical languages such as Greek and Latin are secondary school students – at least 3 million of them across Europe – who experience these languages in their national languages and in isolation from their counterparts across linguistic, if not national, boundaries. We are, however, now poised to develop a new philology, one optimized for a global society, with editions designed to serve many linguistic and cultural communities. Digital technologies provide enabling mechanisms for this new philology but a global philology requires as well a new conception of the relationship between professionalized scholarship and human society, both within Europe and beyond. We have many technological opportunities but the major task before us to assess what sort of scholarly community we wish to support. We have a chance to develop a global republic of letters, one that fosters a true dialogue across civilizations but we must make decisions in the present if we are to pursue changes in the future. This talk goes over possibilities and challenges involved.



---



---

**Wednesday, 4th June**

**SESSION 5 : MANAGING INACCURATE QUOTATIONS**

**Moderator: Smaranda BADILITA (HiSoMA, Lyon)**

**9:00** Introduction. Definitions of the Unstable Notion of Quotation. Brief Survey on the Quotation in the Antiquity (**Smaranda BADILITA**)

Recent work on the concept of quotation in Antiquity ( see, for example C. Nicolas [ed], *Hôs ephat', dixerit quispiam, comme disait l'autre... Mécanismes de la mention et de la citation dans les langues de l'Antiquité*, Recherches & Travaux, hors série, Université Stendhal, ELLUG, Grenoble, 2005 ; C. Darbo-Peschanski [ed], *La citation dans l'Antiquité, Actes du colloque du PARSA Lyon, ENS LSH, 6-8 novembre 2002*, coll. Horos, éd. Jérôme Millon, Grenoble 2004) underline the complexity of this "unstable" notion that affects several areas : the mode of transmission of texts (central issue in ancient studies); their reception (based much more on memory in Antiquity than today); the question of literary genres (some of which have the basic quote : the apophthegms the anthologies, the *memorabilia*); last but not least, social and ideological issues. The term "quotation" does not account for the complexity of this phenomenon in Antiquity, without being accompanied by a "tag cloud": allusion, reference, narrations, *Quellenforschung*, paraphrase, rewrite, intertextuality. Some brief examples will illustrate this multiple shades table.

**9:15** The Typology used in the Biblindex project and its "mapping" with possible encoding patterns (**Laurence MELLERIN and Emmanuelle MORLOCK, HiSoMA**)

This paper discusses methodological issues regarding the analysis of new works from scratch in Biblindex, considering first the nature of a Father's Bible and then describing the way of selecting and characterising a reference to this Bible in a patristic text. These include how to define the difference

between allusion and quotation, how to distinguish and typify the Church Fathers' ways of introducing and changing the biblical texts or how to decide where a quotation begins and where it ends. Through a few case studies, the solutions offered by Biblindex are presented and submitted for discussion.

**9:40** Encoding [Inter]Textual Insertions in Latin "Grammatical Commentary" (**Bruno BUREAU, Christian NICOLAS, Ariane PINCHE, HiSoMA**)

Donatus' Commentary on Terence contains many kinds of inserted texts: quotations from Terence's comedies, and from other Latin and Greek writers, fragments of more or less unknown texts, truncated, abbreviated or inaccurate quotations, mentioned words or phrases. In this paper we intend to show how we have encoded these difficult passages according to TEI guidelines, and to discuss specific difficulties that we encounter now while preparing critical edition of the text.

**10:05** La tradición literaria griega en los ss. III-IV d.C. gramáticos, rétores y sofistas como fuentes de la literatura greco-latina (**Lucía RODRÍGUEZ-NORIEGA GUILLÉN, University of Oviedo**)

**10:20** Dealing with All Kinds of Quotations [and their Parallels] in a Closed Corpus (**Lucía RODRÍGUEZ-NORIEGA GUILLÉN**)

Our project aims to trace and classify all kinds of quotations (literal citations, paraphrases, lax references, imitations, parodies, mere mentions of authors and works), both explicit (with or without mention of the author and/or title) and hidden (imitations, parodies, allusions, use of material without mentioning its origin), in a corpus made up of the Greek grammarians, rhetoricians and "sophists" of the third and fourth centuries AD. At the same time, we try to detect whether or not these are first-hand quotations, and if our authors are, in turn, secondary sources of the same citations in later authors. We also study the philological (textual) aspects of the quotations in their context, and the problems of limits they sometimes pose. Finally, we are interested in the function of the quotation in the citing work. A coordinated project studies quotations in the Latin grammarians of the same period. In my talk I will briefly explain our methodology and how we store all those data in our file cards. In the future we would like to develop a dynamic web-site, but so far, and as a provisional solution, we use a simple desktop application called "Fichas", which has been developed in C# language, under the .Net 4.0 Framework. The application allows the creation and modification of file cards (one for each quotation), which are



stored individually in XML 1.0 format, and eventually exported to PDF format in order to publish them in the (static) website of the Project.

**10:45** Break

**11:00** Text Re-use and Narrative Context: Digital Narratology? (**Lavinia GALLI MILIC, Damien NELIS, Department of Classics, University of Geneva**)

In this paper we want to raise the question of the relationship between on-going work on digital discovery of text re-use and the fact that the corpora of texts that are being searched are made up, at least in part, of narratives of various kinds. As Classicists working mainly from a traditional philological perspective, we focus on Greek and Latin epic poetry. This genre has the advantage of being so strongly codified in terms of its basic narrative features that it lends itself easily to a very simple type of formal narratological analysis. As a result, it also lends itself to an approach that brings into dialogue forms of literary intertextuality involving both specific kinds of text re-use and narrative features that operate on a non-linguistic level. Our question, therefore, is this: to what extent do we need to take into account original narrative contexts and structures when thinking about how to trace, analyse and represent digital analysis of text re-use?

**11:25** Inaccurate Citations as a Source for the ECM of Acts - Misquotations or Lost Manuscript Text? (**Gunnar BÜSCH, INTF**)

The presentation will deal with the methodology used for the ECM of Acts regarding patristic citations that contain text not extant in manuscripts. Special emphasis will be placed on the many differences of using such „inaccurate“ citations as textual witnesses for a critical edition of the New Testament, the main problem being the distinction between unintentional misquotations, deliberate changes by the citing author or even accurate citations of manuscript text unknown to us today. Examples from the ongoing work on the ECM will be used to show how and to what extent distinctions like this can be made.

**11:50** QuotationFinder – Establishing the Degree to Which a Quotation or Allusion Matches Its Source (**LUC HERREN**)

In the first paper on the QuotationFinder software, the tool and its purpose are introduced in a general way. This second paper is more technical and discusses how QuotationFinder produces an ordered list with the most exact quotation first and the loosest verbal parallel last. It is demonstrated how QuotationFinder produces a score for each potential quotation or allusion

by assessing matched words along 5 parameters: quantity (the number of words matched within a reasonable number of lines); quality (whether the exact form of a word is matched, or a different form of the same word, or a form of a cognate); density (how close the matched words are to each other); rarity (how exclusive the matched words are); and, finally, order (to what degree the sequence of the words matches your search text.)

So far, QuotationFinder is aimed at researchers who are looking for quotations of and allusions to relatively small numbers of arbitrary, but short source texts in the vast collections of the TLG (Thesaurus Linguae Graecae) or the CLCLT (CETEDOC Library of Christian Latin Texts.) However, in this presentation we will be looking also at possibilities of extending QuotationFinder in order to facilitate searches of more and longer sources as well as searches in collections other than TLG and CLCLT and in languages other than Latin and Greek.

**12:10 Methodology Used in LOFTS: the Digital Athenaeus Project**  
**(Monica BERTI, University of Leipzig)**

The *Digital Athenaeus* is producing the first digital edition of the *Sophists at Banquet* of Athenaeus of Naucratis. This work is a huge collection of quotations of lost authors and therefore a very interesting use case for implementing ontologies and data models for working with text reuses.

**12:30 Lunch**

---

---

### CONCLUSIONS, OPPORTUNITIES FOR COLLABORATION

**14:00** Towards a Cheatsheet for the Encoding of Biblical References and Quotations (**Emmanuelle MORLOCK**)

The aim of this workshop is to present the work done in the context of the Biblindex project and discuss a draft complementing the existing page on the wiki of [www.tei-c.org](http://www.tei-c.org) ([http://wiki.tei-c.org/index.php/Critical\\_Editions\\_Cheatsheet#Biblical\\_references\\_and\\_quotations](http://wiki.tei-c.org/index.php/Critical_Editions_Cheatsheet#Biblical_references_and_quotations)).

**14:30** Round-table Discussion: Modelling a Network of Quotations.  
Sharing encoding patterns, reference systems and concordances tables.

**Leader: Emmanuelle MORLOCK and Pierre-Édouard PORTIER**

**15:30** Possible EU Projects: H2020, 2. S6, COST, Culture Creative Europe  
(**Émilie SABLON, Lyon Ingénierie Projet**)

**16:30** Discussion/Conclusions

**17:30** End of the meeting

**18:00** end



---



---

## Table of Contributors and/or Participants

### **BADILITA Smaranda** [smaranda.badilita@mom.fr](mailto:smaranda.badilita@mom.fr)

*Smaranda Badilita is a Classicist by training, PhD (2007) in Hellenistic Judaism and Ancient Christianity. She is working since 2011 at the Institut des Sources chrétiennes for the Biblindex Project and is also responsible to review volumes of the collection of "Sources chrétiennes". [Biblindex]*

### **BERLINCOURT Valéry** [valery.berlincourt@unibas.ch](mailto:valery.berlincourt@unibas.ch)

*Valéry Berlincourt (PhD Neuchâtel 2008) is a researcher in Classical Philology at the University of Basel. His current research project, funded by the Swiss National Science Foundation (SNSF "Ambizione" project 148064, years 2014-2016 [<http://p3.snf.ch/project-148064>]), is dedicated to inter- and intra-textuality in the "panegyric epics" of the latin poet Claudian (4th-5th c. CE). Other research activities bear on epic poetry of the Flavian period (1st c. CE), with a focus on Statius, and also on its transmission and reception in the early modern period (with a focus on the 17th c.).<https://klaphil.unibas.ch/latinistik/personen/berlincourt-valery/>[Greek and latin epic poetry]*

### **BERTI Monica** [monica.berti@uni-leipzig.de](mailto:monica.berti@uni-leipzig.de)

*Monica Berti. Classicist and Digital Humanist, currently Academic Assistant at the Humboldt Chair of Digital Humanities at the University of Leipzig. <http://www.monocaberti.it/about/> [LOFT]*

### **BLACKWELL Chris** [christopher.blackwell@furman.edu](mailto:christopher.blackwell@furman.edu)

*Christopher Blackwell is the Louis G. Forgione University Professor of Classics at Furman University, in South Carolina. Since 2001 he has collaborated on digital library architecture with Neel Smith and the editors of the Homer Multitext. He has published two books on Alexander the Great, and has an ongoing collaborative project in historical botany. [CITE]*

**BRUNET Michèle** [michele.brunet@mom.fr](mailto:michele.brunet@mom.fr)

*Michèle Brunet has been since 2006 Professor of Greek Epigraphy at University Lumière – Lyon 2. In 2009, Michèle Brunet has been granted a five-year senior research fellowship from the Institut Universitaire de France for a Digital Humanities Project concerning the Louvre collection of Greek Inscriptions, and in 2012, she got a 3-year grant from the ANR for launching this project. [Epidoc]*

**BÜCHLER Marco** [mbuechler@gcdh.de](mailto:mbuechler@gcdh.de)

*Marco Büchler holds a Diplom in computer science. From April 2008 to March 2011 Marco served as the technical project manager for the eAQUA project and has continued to work in this capacity for yet another project, eTraces, since July 2011. His research interests revolve primarily textual transmissions and related text mining techniques. In addition to his primary responsibilities, Marco manages the Medusa project, as well as the Leipzig Linguistic Services. In March 2013 he completed a PhD in the field of eHumanities. As of May 2014, he is leader of the eTRAP Early Career Research Group. [e-traces]*

**BUREAU Bruno** [brbureau@gmail.com](mailto:brbureau@gmail.com)

*Bruno Bureau is a Professor of Latin language and literature in Jean Moulin University (Lyon). His research interests are Late Latin literature (pagan and Christian) and Latin poetry. In collaboration with Prof. Christian Nicolas (Jean Moulin, Lyon), he created the Hyperdonat project, and gave a digital edition with French translation, critical notes, lexical tools and thesaurus of Donatus' commentary on Terence (<http://hyperdonat.humanum.fr/>). He is now preparing a critical on-line edition of this commentary and a new critical edition of Arator's Historia Apostolica, with French translation and commentary for the Collection des Universités de France (in collaboration with Prof. P.A. Deproost (UCL Louvain). [Hyperdonat]*

**BÜSCH Gunnar** [gunnar.buesch@web.de](mailto:gunnar.buesch@web.de)

*Gunnar Büsch is currently research associate at the INTF in Münster, working on patristic citations for the ECM. He has a PhD-project on Tertullian's Exegesis of Paul. [ECM]*

**CALABRETTO Sylvie** [sylvie.calabretto@insa-lyon.fr](mailto:sylvie.calabretto@insa-lyon.fr)

*Professeur. Enseignante au Département Informatique de l'Institut National des Sciences Appliquées (INSA) de Lyon. Chercheur au Laboratoire d'Informatique en Image et Systèmes d'Information (LIRIS - Equipe DRIM). Directrice de l'Ecole Doctorale InfoMaths de Lyon ([ED InfoMaths](#))*

**CAPRON Laurent** [laurent.capron@vjf.cnrs.fr](mailto:laurent.capron@vjf.cnrs.fr)

*Laurent Capron est docteur de l'université Paris-Sorbonne. Il a été pendant neuf ans ingénieur d'étude à l'Institut de papyrologie de la Sorbonne où il avait, entre autres, la charge de la conservation et de la restauration des collections papyrologiques. Il est maintenant éditeur-rédacteur à L'Année philologique (CNRS – UPR 76). Il a déjà publié plusieurs papyrus grecs issus des collections du musée du Louvre dont il a travaillé à l'inventaire et à la restauration.*

**COFFEE Neil** [ncoffee@buffalo.edu](mailto:ncoffee@buffalo.edu)

*Neil Coffee is Associate Professor of Classics at the University at Buffalo, SUNY. He leads the Tesserae Project, which offers an online computational tool for the study of intertextuality among classical and modern authors. He is author of *The Commerce of War: Exchange and Social Order in Latin Epic* (2009), and is preparing a book on the socioeconomic thought and expression in republican and early imperial Rome. [Tesserae]*

**CRANE Gregory** [gregory.crane@tufts.edu](mailto:gregory.crane@tufts.edu)

*Gregory Crane, Leipzig/Tufts, is interested in opportunities for the study of historical languages in general and Greek and Latin in particular. He is head of the [Perseus Project](#) at Tufts University, holder of [The Humboldt Chair of Digital Humanities at the University of Leipzig](#) which launched the Open Philology Project [Perseus]*

**DIETRICH Marc** [mfvdietrich@gmail.com](mailto:mfvdietch@gmail.com)

*Doctorant contractuel à l'Université Lyon 2 (depuis 2013) : Poétique du récit médical hippocratique (Epidémies I-III ; V-VII) (dir. I. Boehm)*

**ECH CHAEL Yasmine** [yasmine.echchael@mom.fr](mailto:yasmine.echchael@mom.fr)

*Ingénieur de recherche CNRS dans l'équipe des Sources Chrétiennes.*

**EGYED-ZSIGMOND Elöd** [elod.egyed-zsigmond@insa-lyon.fr](mailto:elod.egyed-zsigmond@insa-lyon.fr)

*Elöd Egyed-Zsigmond, researcher at the [LIRIS](#), member of the [DRIM](#) team, is a full time associate professor at the [Computer Science Department](#) of the [INSA-Lyon](#) since 2004. He is working on knowledge management in documents and document manipulation assistance. [Bibindex]*

**FRANZINI Emily** [efranzini@informatik.uni-leipzig.de](mailto:efranzini@informatik.uni-leipzig.de)

*Emily Franzini, a Classicist with an MSc in Management Science & Innovation, is currently a Research Associate at the Humboldt Chair for Digital Humanities at the University of Leipzig. Her interests lie in Multilingualism and searching for ways in which to make the Humanities commercially appealing. [e-traces]*

**FRANZINI Greta** [franzini@informatik.uni-leipzig.de](mailto:franzini@informatik.uni-leipzig.de)

*Greta Franzini is a Classicist by training and has an MA in Digital Humanities. She is currently a part time Digital Humanities PhD student at the UCL Centre for Digital Humanities and works as a full time Research Associate for the Humboldt Chair of Digital Humanities at the University of Leipzig. Her research interests revolve around the fields of Electronic Editing, Paleography and Manuscripts Studies. [e-traces]*

**GALLI Milic** [Lavinia.Lavinia.GalliMilic@unige.ch](mailto:Lavinia.Lavinia.GalliMilic@unige.ch)

*Lavinia Galli Milic, Université de Genève, travaille actuellement sur le poète épique Valérius Flaccus dans le cadre d'un projet du Fonds national Suisse pour la recherche scientifique intitulé "Intertextuality in Flavian Epic Poetry ». Le projet prévoit la collaboration avec l'équipe de Buffalo qui a développé l'outil Tesserae (plus précisément avec Neil Coffee et Chris Forstall). Voir <http://www.unige.ch/lettres/antic/latin/enseignants/LaviniaGALLIMILIC.html>. [Greek and latin epic poetry]*

**GESCHE Samuel** [samuel.gesche@gmail.com](mailto:samuel.gesche@gmail.com)

*Samuel Gesche was a post-doctoral researcher from the LIRIS lab. He worked on the aspect of quotation finding, and continued working on this project since then. [Bibindex]*

**GOUDARD David** (HiSoMA) [david@goudard.web4me.fr](mailto:david@goudard.web4me.fr)

*Informaticien en charge du développement de Bibindex.*

**GRIGNON Sébastien** [sgrignon@vjf.cnrs.fr](mailto:sgrignon@vjf.cnrs.fr)

*Co-directeur de l'Année Philologique*

**GUÉRARD Marie-Gabrielle** (HiSoMA) [marie-gabrielle.guerard@mom.fr](mailto:marie-gabrielle.guerard@mom.fr)

*Ingénieur de recherche CNRS dans l'équipe des Sources Chrétiennes.*

**HERREN Luc** [lherren@uni-muenster.de](mailto:lherren@uni-muenster.de)

*Luc Herren, University of Münster, studied theology in Bern, Harvard, and Heidelberg. He started programming computers because he needed to find quotations and allusions for a dissertation project (the reception history of a biblical text, Magnificat, Lk 1:46-55). He is an assistant editor of the Nestle-Aland Novum Testamentum Graece at the INTF in Münster and currently lives in London. [Quotation Finder]*



**INGARAO Maud** [maud.ingarao@ens-lyon.fr](mailto:maud.ingarao@ens-lyon.fr)

*Institut d'Histoire de la pensee classique UMR 5037 - IHPC*

**KALVESMAKI Joel** [KalvesmakiJ@doaks.org](mailto:KalvesmakiJ@doaks.org)

*Joel Kalvesmaki, Dumbarton Oaks (Harvard University). He is the Editor in Byzantine Studies, where he manages print and digital publications. His research focuses on Greek patristics and symbolism in antiquity. The synthesis of his scholarship and digital humanities work is best exemplified by the online reference work Guide to Evagrius Ponticus, evagriusponticus.net.[The Text Alignment Protocol]*

**KEPEKLIAN Gabriel** [gabriel.kepeklian@wanadoo.fr](mailto:gabriel.kepeklian@wanadoo.fr)

*As research engineer at Thomson (1981-1988) in the field of computer languages and formal grammars, Gabriel Kepeklian - INSA 1981 (National Institute of Applied Sciences of Lyon) - has developed compiler compilers and created supercomputer control languages. Later, he has been system architect and product manager. He is currently responsible for R & D at Atos. His projects are focused on the technologies of web of data and linked data. Gabriel Kepeklian is also Chairman of the DataLift association, which promote the development of the web of data, research, innovation and any activity to promote its usages. Nearby, he studies ancient Armenian (texts translation, grammar). Web: <http://www.kepeklian.com/blog> [Datalift]*

**KRÜGER Volker** [volker.krueger@uni-muenster.de](mailto:volker.krueger@uni-muenster.de)

*Volker Krüger is Research associate at the Institute for New Testament Textual research. [ECM]*

**MACLACHLAN Rosalind** [R.F.MacLachlan@bham.ac.uk](mailto:R.F.MacLachlan@bham.ac.uk)

*Rosalind MacLachlan, Research Fellow on the COMPAUL project looking at the earliest commentaries on the Pauline Letters in the Institute for Textual Scholarship & Electronic Editing (ITSEE) at the University of Birmingham, UK. Previously, Research Fellow on the Vetus Latina Iohannes Project producing a edition of the Old Latin of the Gospel of John, also at ITSEE. She is a Classicist by training. [COMPAUL]*

**MELLERIN Laurence** [laurence.mellerin@mom.fr](mailto:laurence.mellerin@mom.fr)

*Laurence Mellerin, research engineer at the Institut des Sources Chrétiennes, has been in charge of the Biblindex project since 2006. She is a Classicist and Theologian by training, with a strong focus on digital humanities. [Biblindex]*

**MORITZ Maria** [moritz@informatik.uni-leipzig.de](mailto:moritz@informatik.uni-leipzig.de)

*Maria Moritz, Computer Scientist at the Humboldt Chair for Digital Humanities University of Leipzig, where she works on data preparation and application development. Her main interests are Natural Language and Text Processing and Investigation. [e-traces]*

**MORLOCK Emmanuelle** [emmanuelle.morlock@mom.fr](mailto:emmanuelle.morlock@mom.fr)

*Emmanuelle Morlock currently works as a research officer at the CNRS in the HiSoMA laboratory in Lyon. Her main mission is to assist researchers in their use of digital humanities tools and methodologies in the areas of digital databases and scholarly digital editions. [Biblindex]*

**NELIS Damien** [Patrick.Damien.Nelis@unige.ch](mailto:Patrick.Damien.Nelis@unige.ch)

*Damien Nelis is Professor of Latin in the University of Geneva. Before that he taught in the University of Durham, UK, and in Trinity College Dublin. He works mainly in the field of Latin epic poetry and is currently involved in a research project on digital humanities in a collaboration with the TESSERA project involving Neil Coffee, Chris Forstall and Lavinia Galli Milic and funded by the Swiss National Science Foundation. [Greek and latin epic poetry]*

**NICOLAS Christian** [christian.nicolas@univ-lyon3.fr](mailto:christian.nicolas@univ-lyon3.fr)

*Christian Nicolas is a Professor of Latin language and literature in Jean Moulin University (Lyon). In collaboration with Prof. Bruno Bureau (Jean Moulin, Lyon), he created the Hyperdonat project, and gave a digital edition with French translation, critical notes, lexical tools and thesaurus of Donatus'commentary on Terence (<http://hyperdonat.huma-num.fr/>). [Hyperdonat].*

**PAPARNAKIS Athanasios** [apaparna@past.auth.gr](mailto:apaparna@past.auth.gr)

*Athanasios Paparnakis, Associate Professor, Faculty of Theology, Aristotle University of Thessaloniki, is a biblical scholar working on the Old Testament and Biblical Theology with special emphasis on the reception history of the Bible in the patristic tradition. He has been working on data processing and databases of biblical quotations in the patristic texts. [Digital Greek Catena]*

**PATARIDZE Tamar** [tamar.pataridze@uclouvain.be](mailto:tamar.pataridze@uclouvain.be)

*Tamar Pataridze, Philologist, Teacher of Georgian Language and Literature; PhD in Old Georgian literature / Georgian-Byzantine literary contacts (I. Javakhishvili **Tbilisi State University**, Georgia); PhD in Languages and Literature (Institute of Oriental Studies, Université Catholique de*

Louvain, Belgium). Current Occupation: « Chargé de Recherches », F.N.R.S (Université Catholique de Louvain, Belgium). [GREgORI]

**PINCHE Ariane** [ariane.pinche@gmail.com](mailto:ariane.pinche@gmail.com)

*Ariane Pinche is currently trainee for the Hyperdonat Project as part of her training course at the Ecole des Chartes (second year of MA : « Digital Technologies applied to History »). She has a MA in classical literature. Her research interests are medieval literature, old French and digital editions. She is working now on digital and technical aspects of the Hyperdonat Project. She is setting up the TEI and XSL files in sight of a digital edition of the corpus. [Hyperdonat]*

**PORTIER Pierre-Édouard** [pierre.edouard.portier@gmail.com](mailto:pierre.edouard.portier@gmail.com)

*Pierre-Édouard Portier, researcher at the [LIRIS](#), is a full time associate professor at the [Computer Science Department](#) of the [INSA-Lyon](#). His main working files are the use of Web data to improve information retrieval and the construction of complex documents in specialized editorial contexts, applied to digital humanities. [Bibindex]*

**PSILAKIS Catherine** [ca\\_psilak@yahoo.fr](mailto:ca_psilak@yahoo.fr)

*[Université Lille 3](#), [CNRS-UMR 8163 "Savoirs, Textes, Langage" \(STL\)](#), Graduate Student*

**RODRÍGUEZ-NORIEGA Guillén** [Lucía.lnoriega@uniovi.es](mailto:Lucía.lnoriega@uniovi.es)

*Lucía Rodríguez-Noriega Guillén, University of Oviedo (Spain). Her main fields of research are textual criticism (for instance, she has edited a fragmentary author, Epicharmus, and co-edited Aelian's *Natura Animalium* for the Teubner's series), fragmentary comedy and its language (besides her publications and the supervision of several short and PDh thesis on the subject, she has headed two Projects on the language of Greek fragmentary comic writers of the 5<sup>th</sup> and 4<sup>th</sup> centuries, funded by the Spanish Ministry of Economy and Competitiveness), and Athenaeus of Naucratis (besides other publications, she is currently working on the first complete Spanish annotated translation of the *Deipnosophists*, of which she has already published 4 volumes, a fifth one being now in press). Currently she heads the Project *La tradición literaria griega en los ss. III-IV d.C. gramáticos, rétores y sofistas como fuentes de la literatura greco-latina*, also funded by the Spanish Ministry of Economy and Competitiveness. (For her complete CV see <http://www.lnoriega.es/curriculum.html>). [La tradición literaria griega]*

**ROUECHÉ Charlotte** [charlotte.roueche@kcl.ac.uk](mailto:charlotte.roueche@kcl.ac.uk)

*Charlotte Roueché, Senior Research Fellow in Digital Hellenic Studies, King's College London. Research interests: epigraphy, Byzantine studies, digital humanities. [SAW]*

**SABLON Émilie** [Emilie.Sablon@lip-lyon1.fr](mailto:Emilie.Sablon@lip-lyon1.fr)

*Chargée de mission à Lyon Ingénierie Projet.*

**SCHULTHESS Sara** [sara.schulthess@unil.ch](mailto:sara.schulthess@unil.ch)

*Sara Schulthess is a Swiss National Science Foundation PhD student at the University of Lausanne. She works on the Arabic manuscripts of the Pauline letters: listing of the manuscripts, edition of manuscript Vat. Ar. 13 and analysis of the role of the digital for the research field. [Institut romand des sciences bibliques]*

**SMITH Catherine** [c.j.smith@bham.ac.uk](mailto:c.j.smith@bham.ac.uk)

*Catherine Smith, Technical officer on the COMPAUL project and Workspace for Collaborative editing in the Institute for Textual Scholarship & Electronic Editing (ITSEE) at the University of Birmingham, UK. Her background is in theology, linguistics and computer science. [COMPAUL]*

**SMITH Neel** [nsmith@holycross.edu](mailto:nsmith@holycross.edu)

*Associate Professor (Ph.D., University of California, Berkeley). Fields: Classical archaeology, ancient geography, implications of new information technologies for Classical studies. [CITE]*

**VAN Elverdinghe Emmanuel** [emmanuel.vanelverdinghe@uclouvain.be](mailto:emmanuel.vanelverdinghe@uclouvain.be)

*Emmanuel Van Elverdinghe, Philologist, MA in Classical Studies is currently a Research Fellow of the Belgian Fund for Scientific Research (“aspirant F.R.S.-FNRS”). His researches take place at the Université catholique de Louvain (Louvain-la-Neuve), and concentrate on Armenian manuscripts. [GREORI]*

**VAN PEURSEN Wido** [w.t.van.peursen@vu.nl](mailto:w.t.van.peursen@vu.nl)

*Wido van Peursen, Eep Talstra Centre for Bible and Computer (ETCBC), VU University Amsterdam, is Professor of Old Testament, with a strong focus on Digital Humanities. He has been involved in various cooperation projects of the Peshitta Institute Leiden (PIL) and the ETCBC, till 2012 as member of the PIL and since September 2012 as head of the ETCBC. [SHEBANQ]*

**WETZIG Norman** [normanwetzig@gmx.de](mailto:normanwetzig@gmx.de)

*Norman Wetzig, Gradstudent at the Rheinische-Friedrichs-Wilhelm-Universität Bonn [CASG]*

---



---

## List of Projects

### **BIBLINDEX (HISOMA AND LIRIS, LYON)**

<http://www.bibindex.org>

<http://bibindex.hypotheses.org/>

**Session 1:** The Biblindex Project, Index of Biblical Re-Uses in the Early Christian Literature

**Session 2:** Was it better before? Automated Quotation Detection in Ancient Texts

**Session 3 :** Introduction

**Session 4:** Tei Encoding Patterns for Citations, Quotations and Allusions: What's "in Stock" ?

**Session 5:** Introduction. Definitions of the Unstable Notion of Quotation. Brief Survey on the Quotation in the Antiquity

**Session 5:** The Typology Used in the Biblindex Project and its "Mapping" with Possible Encoding Patterns

**Session 5:** Towards a Cheatsheet for the Encoding of Biblical References and Quotations

**Session 5:** Round-table Discussion: Modelling a Network of Quotations

### **CASG. CORPUS DER ARABISCHEN UND SYRISCHEN GNOMOLOGIEN**

<http://casg.orientphil.uni-halle.de/?lang=en>

**Session 1:** The "Corpus of Arabic and Syriac Gnomologia" (CASG)

### **THE CITE ARCHITECTURE (Holy Cross College, Furman University)**

<http://cite-architecture.github.io>

**Session 4:** Using CTS for Hierarchical Annotations

### **COMPAUL (ITSEE, BIRMINGHAM)**

<http://arts-itsee.bham.ac.uk/itseeweb/epistulae/compaul.htm>

**Session 1:** The Vetus Latina Iohannes & COMPAUL Projects

**DATALIFT (ANR Contint, France)**

<http://datalift.org/>

**Session 3** : Utilisation de la plateforme d'élévation de données DataLift dans un cadre linguistique, application à l'arménien classique

**DIGITAL GREEK PATRISTIC CATENA**

<http://dgpc.web.auth.gr/index.php>

**Session 1** : Digital Greek Patristic Catena

**ECM. EDITIO CRITICA MAIOR (INTF, Münster)**

<http://www.uni-muenster.de/INTF/ECM.html>

**Session 1**: The Digital Processing of Patristic Citations for the Editio Critica Maior (ECM) of Acts

**Session 5**: Inaccurate Citations as a Source for the ECM of Acts –Misquotations or Lost Manuscript Text?

**E-PIGRAMME (HISOMA, Lyon)**

[http://www.hisoma.mom.fr/mb/IG\\_LOUVRE/E-PIGRAMME-FR.html](http://www.hisoma.mom.fr/mb/IG_LOUVRE/E-PIGRAMME-FR.html)

**Session 4**: Visual Quotation

**E-TRACES (GCDH, Göttingen)**

<http://etraces.e-humanities.net/>

**Session 1** : eTRACES, eTRAP

**Session 2**: Introduction: Overview of the Possible Methodologies Used in Text Retrieval

**Session 2**: Round-Table Discussion

**Session 4**: Introduction: Requirements for a Digital Ecosystem

**Session 4**: Round-Table Discussion: Is a Common Methodology Possible?

**GREEK AND LATIN EPIC POETRY (Department of Classics, University of Geneva)**

<http://www.unige.ch/lettres/antic/index.html>

**Session 5**: Text Re-Use and Narrative Context: Digital Narratology?

**GREGORI: SOFTWARES, LINGUISTIC DATA AND CORPUS FOR ANCIENT GREEK AND ORIENTAL LANGUAGES (CIOL, University of Louvain)**

<http://www.uclouvain.be/ciol.html>

**Session 3**: GREgORI: Softwares, Linguistic Data and Corpus for Ancient Greek and Oriental Languages

**HYPERDONAT (HiSOMA, Lyon)**

<http://hyperdonat.huma-num.fr/>

**Session 5:** Encoding (Inter)Textual Insertions in Latin "Grammatical Commentary"

**INSTITUT ROMAND DES SCIENCES BIBLIQUES (Lausanne)**

<http://www.unil.ch/irsb>

**Session 3 :** Round-Table Discussion

**LOFTS. LEIPZIG OPEN FRAGMENTARY TEXTS SERIES (University of Leipzig)**

<http://www.dh.uni-leipzig.de/wo/projects/open-greek-and-latin-project/the-leipzig-open-fragmentary-texts-series-lofts/>

**Session 1 :** The Leipzig Open Fragmentary Texts Series (LOFTS)

**Session 4:** Methodology for LOFTS Fragments: the *Digital Fragmenta Historiarum Graecorum* Project and the *Digital Marmor Parium* Project (Monica Berti, University of Leipzig)

<http://www.dh.uni-leipzig.de/wo/open-philology-project/the-leipzig-open-fragmentary-texts-series-lofts/digital-fragmenta-historiarum-graecorum-dfhg-project/>

**Session 5:** Methodology Used in LOFTS: the Digital Athenaeus Project

<http://www.dh.uni-leipzig.de/wo/projects/open-greek-and-latin-project/digital-athenaeus/>

**PERSEUS DIGITAL LIBRARY**

<http://www.perseus.tufts.edu/hopper/>

**Lecture:** June 1914 and Philology for the 21<sup>st</sup> Century

**QUOTATIONFINDER (Münster, London)**

<http://quotationfinder.uni-muenster.de/>

**Session 2:** QuotationFinder – Searching for Quotations and Allusions in Greek and Latin Texts

**Session 5:** QuotationFinder – Establishing the Degree to Which a Quotation or Allusion Matches Its Source (Luc Herren)

**SAW. SHARING ANCIENT WISDOMS (King's College, London)**

<http://www.ancientwisdoms.ac.uk/>

**Session 1:** The Sharing Ancient Wisdoms Project: Aims, Problems and Achievements

**SHEBANQ (ETCBC, Amsterdam)**

<http://www.godgeleerdheid.vu.nl/nl/onderzoek/instituten-en-centra/eep-talstra-centre-for-bible-and-computer/index.asp>

**Session 3:** Exploring New Directions in the Computational Analysis of Syriac Texts

**TESSERAЕ (University at Buffalo)**

<http://tesserae.caset.buffalo.edu/>

**Session 2:** Modeling the Scholars: Detecting Intertextuality through Enhanced Word-Level N-Gram Matching

**THE TEXT ALIGNMENT PROTOCOL (Dumberton Oaks)**

<http://evagriusponticus.net/>

**Session 3:** The Text Alignment Protocol, a Proposed Data Model for Interchangeable Bitexts

**Session 4:** Declaring Quotations through Canonical Reference Numbers in the Text Alignment Protocol

**LA TRADICIÓN LITERARIA GRIEGA (University of Oviedo)**

<http://lnoriega.es/Fuentes.html>

**Session 5:** La tradición literaria griega en los ss. III-IV d.C. gramáticos, rétores y sofistas como fuentes de la literatura greco-latina

**Session 5:** Dealing with all kinds of quotations (and their parallels) in a closed corpus: the methodology of the project: “The Literary Tradition in the Third and Fourth Centuries AD: Grammarians, Rhetoricians and Sophists as Sources of Graeco-Roman Literature”.