

Open Science für die Geisteswissenschaften  
Aus den Erfahrungen in Deutschland

# 人文学のためのオープンサイエンス ドイツでの体験から

So Minagawa

宮川 創

筑波大学 人文社会系 准教授

# 宮川創のプロフィール

- 北海道大学(学士)→ 京都大学(修士)→ ゲッティンゲン大学(博士)→関西大学(ポスドク)→京都大学(助教)→国語研(助教)→筑波大(准教授)
- ゲッティンゲン大学エジプト学・コプト学専修でDr. phil. (Koptologie)取得
- **古代末期エジプト**で、エフェソス公会議とカルケドン公会議という、キリスト教史上、重要な時期を生きた**シェヌーテの手紙における詩篇の引用**を中心に研究(コプト語文献とギリシア語文献を大量に扱う)



東海大学古代エジプト及び中近東コレクション  
ション(東海大学文明研究所)での調査のようす

# ゲッティンゲン大学でコプト学とコーパス言語学とDHを学ぶ

- **ドイツ研究振興協会**(DFG)の5年間のプロジェクトで、コプト語文献のデジタルエディションおよびデジタルコーパスを作成する有給研究員として働きながら修学
- 文献のデジタル化の恩恵を受ける側であると同時に、**文献のデジタル化をする側**でもあった([Coptic SCRIPTORIUM](#)、[Papyri.info](#)、[CoptOT](#)など)

●デジタルコーパスを用いたデータ駆動型の間テキスト性研究：古代末期エジプトの二人の修道院長のコプト語書簡におけるコプト語訳聖書からの引用の検知と分析  
[2021-CH-126 (2021/5/22)] (人文科学とコンピュータ研究会)



客川 健 君 (正会員)

2013年 北海道大学文学部言語・文学コース 卒業。  
2015年 京都大学大学院文学研究科 博士前期課程 修了。  
2015年 日本学術振興会特別研究員 (DC1)。  
2015年 ドイツ研究振興協会 (DFG) 特別研究領域 (SFB) 1136 研究員。  
2015年 ドイツ研究振興協会 (DFG) 全米人文科学基金 (NEH) ジョイントプロジェクト KELLIA 研究員。  
2020年 関西大学アジア・オープン・リサーチセンター PDフェロー。  
2021年 京都大学大学院文学研究科 博士後期課程 研究認定フェロー。  
2021年 京都大学大学院文学研究科 助教。  
2022年 ケールク・アウグスト大学ゲッティンゲン (ゲッティンゲン大学) 博士課程修了 Dr.phil.  
2022年 人間文化研究機構 国立国語研究所 研究系 助教。  
コーパス言語学、デジタルヒューマニティーズ、言語資源アーカイブに関する研究に従事。

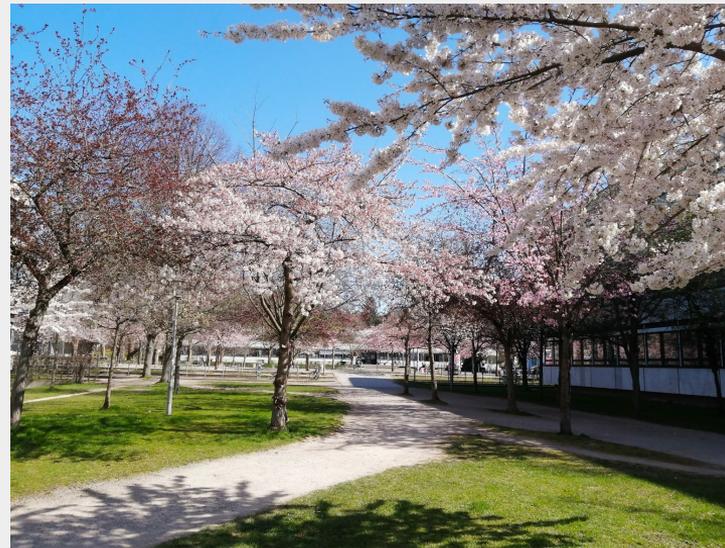
**【推薦理由】** 本論文は、史料のデジタルコーパス化と引用の分析を通してデータ駆動型の間テキスト性研究の可能性を示したものである。候補者は、5世紀エジプトのキリスト教修道院長2人が記したコプト語キリスト教書簡と呼ばれる大量の書簡を、TEI (Text Encoding Initiative) を用いてデジタル翻訳した。その成果に対して形態素解析・係り受け解析を適用することでデジタルコーパス化し、古代のコプト語訳聖書からの引用を検出・探知することで、従来の手動の研究では見つけられなかった聖書からの引用を多数発見し、分析することに成功した。人文科学とコンピュータ研究会の一大トピックである「史料とデジタル技術の新領域を切り開く研究」を優れた形で実践したものであり、高く評価できる。

[戻る]

2022年度情報処理学会  
**山下記念研究賞**  
(「デジタルコーパスを用いたデータ駆動型の間テキスト性研究」)

# ゲッティンゲンとゲッティンゲン大学

- ニーダーザクセン州・最南部のゲッティンゲン郡にある都市
- 1737年創立の  
**Georg-August-Universität  
Göttingen**を中心にした大学都市
- 発表者は、Philosophische Fakultätの  
Seminar für Ägyptologie und  
KoptologieのPromotionに所属
- 同時にDFGの**SFB 1136**で  
Wissenschaftlicher Mitarbeiterとして働く
- 最初の2年間は、DFGとNEH共同プロジェクトKELLIAでもWissenschaftlicher  
Mitarbeiterとして働く



GEORG-AUGUST-UNIVERSITÄT  
GÖTTINGEN

# ドイツの博士課程

- 学費(実質)無料
  - Semesterticketの料金
  - 保険など
- 指導教官が許せば永遠に博士課程に在籍できる
- 福利厚生(州内の電車が無料、市内交通無料、様々な割引など)
- 何らかの形で働いて給料をもらうのが普通
  - 研究員(プロジェクトで働いたり、授業をしたり)
  - 研究助手
  - 図書館や事務の仕事
- 奨学金も充実
  - DAAD
  - Brot für Welt
  - KAAD など



# 研究員(Wissenschaftlicher Mitarbeiter = Research Fellow)

- ドイツの大学は公立大学が多く、公務員とおなじ給料スケール(TV Lスケール)
- 私の場合、最初の2年間は、2つのプロジェクト掛け持ちでTV L 13の95%で、月に3,600ユーロくらいの月収
- 税金が高く、40%くらい取られて、結局2,400ユーロくらい
- ただ、社会保険や失業保険や健康保険が込みで、医療費無料、失業手当が出る、必要なら生活保護 (Harz-IV) も
- 病院での診察や手術や入院は無料、医療レベルは高いと言われている



# 研究員の財源は豊富

- 通常は、講座・専修の資金で雇われるか、プロジェクト資金で雇われる
- プロジェクトの財源
  - 大学・州・市による助成
  - [欧州研究委員会 \(ERC\)](#): 様々な助成
  - [連邦教育・研究省 \(BMBF\)](#): 様々な助成
  - [ドイツ研究振興協会 \(DFG\)](#)
    - Excellenzcluster(5年ごとに選ばれる)
    - SFB 特別研究領域/共同研究センター(4年ごとに更新、最長12年)
    - 他にも多数
  - デジタルインフラストラクチャ系
    - DARIAH-DE, CLARIN-D, CLARIAH, OpenAIRE, NFDI など
  - 学術アカデミー
  - その他、財団系
    - フンボルト財団、フォルクスヴァーゲン財団、フリッツ・テュッセン財団など
- 宮川が雇われたのは、[DFGのSFB \(特別研究領域/共同研究センター\)1136](#)

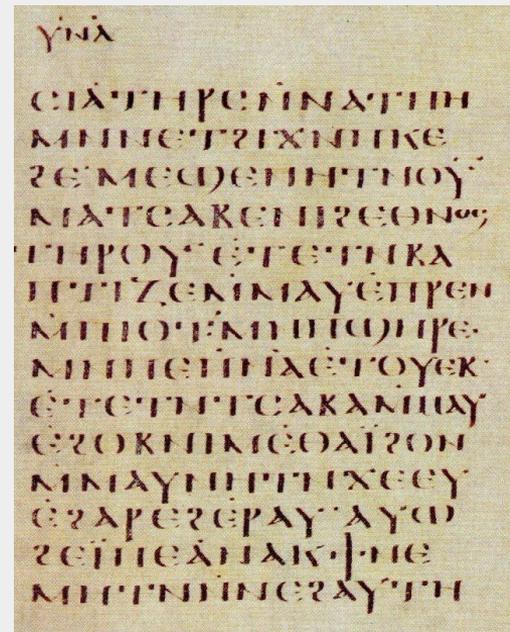
# DFGとSFB 1136

- [Deutsche Forschungsgemeinschaft](#) 「ドイツ研究振興協会」
  - SFB、Exzellenzcluster、Forschergruppeなどの助成を行う
- SFB: Sonderforschungsbereich
  - 日本語では「特別研究領域」？しかし、公式英訳は「Collaborative Research Centre」で、日本語訳は「共同研究センター」？
  - 財政:4年間で6,700,000ユーロ
- 私が働いていたのは、ゲッティンゲン大学に設置されたSFB 1136: **Bildung und Religion in Kulturen des Mittelmeerraums und seiner Umwelt von der Antike bis zum Mittelalter und zum Klassischen Islam** 「古代から中世及び古典イスラム期にかけての地中海圏とその周辺の文化における教育と宗教」



# コプト語は(古代) エジプト語の最終段階

- **アフロ・アジア語族**エジプト語派の言語
- コプト文字で書かれたエジプト語
  - **コプト文字**:ギリシア文字24文字+6-9文字のエジプト民衆文字由来の文字
  - 紀元前のプトレマイオス朝期から、ギリシア文字によるエジプト語表記の試み(Old Coptic)
  - 2-3世紀頃に綴りの標準化が進み、以降コプト文字で書かれたエジプト語をコプト語と呼ぶ
- **貴重な宗教文献**(キリスト教、マニ教、グノーシス主義等)

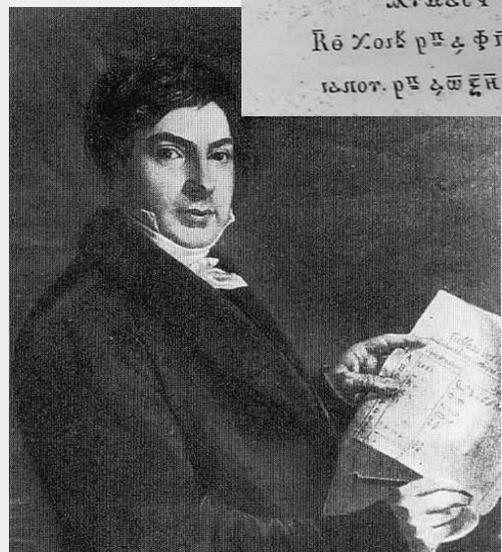


# コプト語の現在と復興運動

- 17世紀にコプト語話者がいたとする記録はあるのを最後に、話し言葉としてコプト語を使う者は確認されていなかったが、19世紀後半から、コプト語を日常言語として復興する**コプト語復興運動**が始まる
- 文語(コプト正教会の典礼言語)としては、コプト語は現代まで使い続けられている
- シャンポリオン以来、古代エジプト語(ヒエログリフ、ヒエラティック、デモティック)資料の**解読の鍵**となる



Κλαυδιος Γωε. λαλην  
Αγυαση Σεν  
Κο Χουε ρε φπζ μ. ετε  
καπορ. ρε φπζ μ. ρε.



# ゲッティンゲン大学エジプト学コプト学専修とコプト語

- シャンポリオンによる「ヒエログリフの解読」(1822)と近代エジプト学の始まり
- **ドイツで2番目にできたエジプト学専修**
- エジプト語はその末裔であるコプト語を用いて解読されてきたことから、当初はコプト語文献学をメインとするコプト学も専修名に加わる
- ドイツでエジプト学コプト学を冠している大学は、ミュンスター大学、ゲッティンゲン大学、ミュンヘン大学だが、コプト学を学べる大学は、ライプチヒ大学、ベルリン自由大学、フンボルト大学、ハイデルベルク大学、ヴュルツブルク大学など数多い



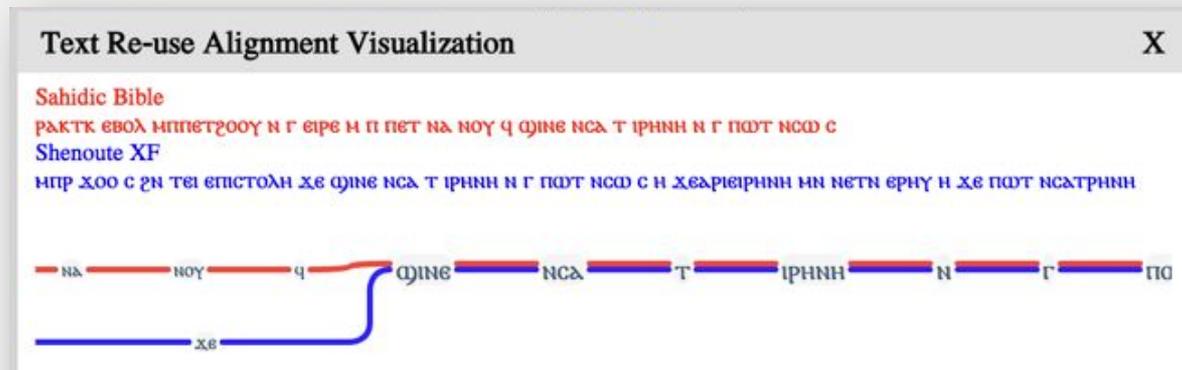
# SFB 1136 1年目 (2015/6-2016/5)

- 宮川は2015/10からSFB 1136に加入
- BMBFのeTRAPプロジェクトとテキスト・リユース(引用や引喩など)探知ソフトTRACERをコプト語に適用させ始める
- コプト語文献の効率的なデジタルコーパス化のためにコプト語OCRの開発を始める
  - eTRAPの主任研究者のMarco Büchlerの情報学演習の授業でこのプロジェクトが開始
  - 情報学の学部生(9年目)のKirill Bulertとともにコプト語OCRモデルの開発に取り組む
- 全米人文科学基金・ドイツ研究振興協会ジョイントプロジェクトであるKELLIAにて、ヨーロッパでのDHにおけるテキストマークアップの基本であるTEIの伝授を受け、TEIを用いて、様々なデータを機械可読化にする

# BMBFプロジェクト eTRAP との共同研究

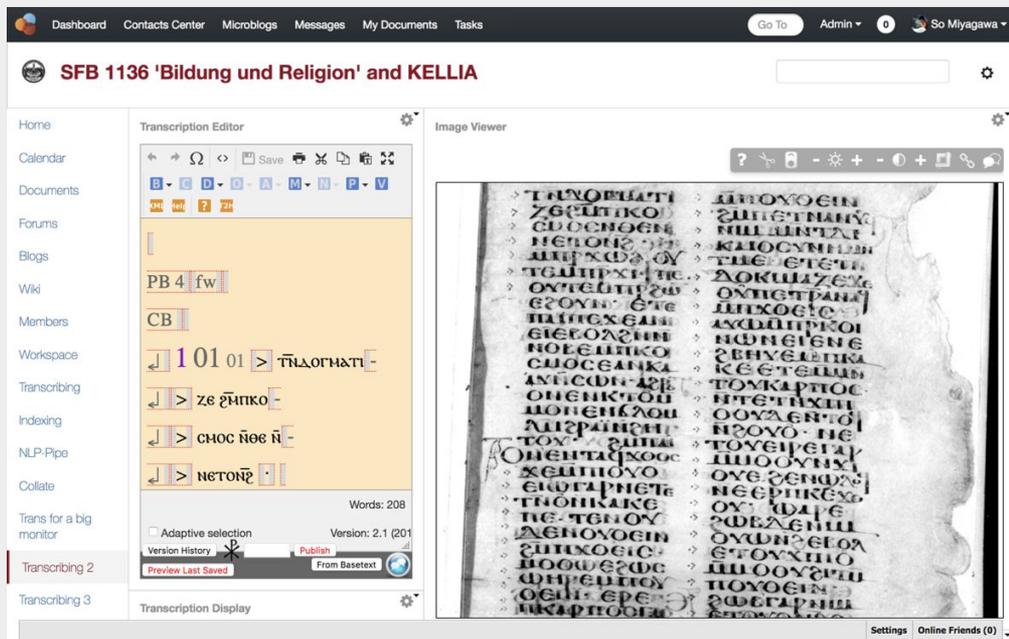
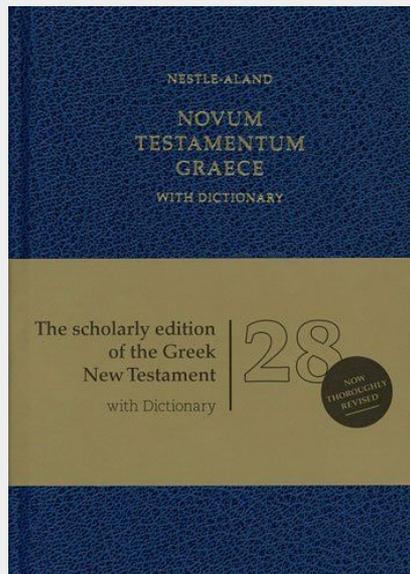


- Bundesministerium für Bildung und Forschung - **BMBF 連邦教育・研究省**のEarly Career Research Group (2015年3月1日から4年間)、1,600,000ユーロ
- **TRACER**というテキスト・リユース(引用や引喩など)探知のプログラムを開発
- 筆者が写本を翻刻して作成した、4-5世紀の修道院長シェヌーテのコプト語コーパスと、ゲッティンゲン科学アカデミーのコプト語訳聖書コーパスの間で、テキスト・リユースを探知



# そのほか、ミュンスター大学新約聖書本文研究所との共同

- ネストレ=アールアント版の拡大版であるEditio Critica Majorの編纂で使われているDigital Manuscript Roomを用いてシェヌーテ書簡のデジタル学術編集版(デジタル・エディション)を作成



# SFB 1136 2年目 (2016/10-2017/9)

- 国際コプト学会(クレアモント、カリフォルニア)にてシヌーテの書簡における初期のTRACERの結果と聖書引用の分析を発表
- この学会前にあったKELLIAの研究集会で、KELLIAホームページの進捗状況と、OCRサイトの進捗状況を発表
- SFBの旅費(1年間2000ユーロ)で、ギリシア、ポーランド、イタリアの学会など行って、発表したり、チュートリアルを受けたりする
- DHの学会: DATeCH2017 (ゲッティンゲン)、AIUCD(ヴェネツィア)、DH2017(モントリオール)



# コプト語OCR

現在は、専ら機械学習系手書きテキスト認識ソフト  
[Transkribus](#)を学習させて用いている

- 2015-2016年当時、ニューラルネットワーク機械学習系OCRの主流だったOCRopusとTesseractをコプト語文献で学習させた

OXFORD  
ACADEMIC

Journals

Books



**DSH** Digital  
Scholarship  
in the  
Humanities

eadh  
EUROPEAN ASSOCIATION  
FOR DIGITAL HUMANITIES

**AD**  
**HO** ALLIANCE OF  
DIGITAL  
HUMANITIES  
ORGANIZATIONS

Issues

Advance articles

Submit ▼

Purchase

Alerts

About ▼

Digital Scholarship in the Humanities



Advanced  
Search



Volume 34, Issue  
Supplement\_1  
December 2019

JOURNAL ARTICLE

## Optical character recognition of typeset Coptic text with neural networks <sup>FREE</sup>

So Miyagawa ✉, Kirill Bulert, Marco Büchler, Heike Behlmer

*Digital Scholarship in the Humanities*, Volume 34, Issue Supplement\_1, December 2019, Pages i135–i141, <https://doi.org/10.1093/dsh/llc/fqz023>

**Published:** 22 April 2019



PDF

Split View

Cite

Permissions

Share ▼

VIEWS



More metrics information

ALTMETRIC



Email alerts

# SFB 1136 3-4年目 (2017/10 – 2019/9)

- 博論の執筆を本格化させる
- アメリカ合衆国、スウェーデン、イギリス、フランス、イタリア、イスラエルなどに研究調査滞在
- フランス国立図書館やナポリ国立図書館や大英図書館などの写本のデジタル翻刻も完成させる
- 第二期に向けて、博論を執筆するとともに、ポスドクになる準備もする
- 指導教官との幾度にも渡る執筆・コメント・修正の繰り返しで大体博論はでき、第二期の Wissenschaftlicher Mitarbeiterの内定ももらった
- DFGのEvaluationは良い結果だったらしく、第二期は確実だと思われた
- →しかし、二期目が始まる1ヶ月前にSFB1136は、DFGの財政的な理由から(?)第二期目を続行できないことに
- →Arbeitsamtで急いで失業手当の申請をする
- しかし、失業手当が出る5日前に、DFGが、既に二期目に内定している人限定で、1年間雇用する通知が来る

# SFB1136 5年目 (2019/6-2020/5)

- 第二期目に内定していた人のみがSFBに残る
- 残った人も別の内定を受け、次々と去っていく
- 博士論文を一旦はほぼ完成させ、スウェーデン・ウプサラ大学のエジプト学の学会で一部を発表するが、2020/6からのことを案じて不安になる
- 大英博物館のコプト語写本のカタログ作成(9ヶ月)の職に受かるが、ブレクジットによって、労働条件など色々変わって行けないことになった
- ヘブライ大学の1年間の留学プログラムに採択される
- 3月から本格的にコロナが蔓延しているときに、関西大学から職のオファーを受ける
- 指導教官の薦めにしたがって、2020/4から関西大学の[アジア・オープン・リサーチセンター](#)「KU-ORCAS」で働き始める



# コプト語での自然言語処理とコーパス

## COPTIC SCRIPTORIUM

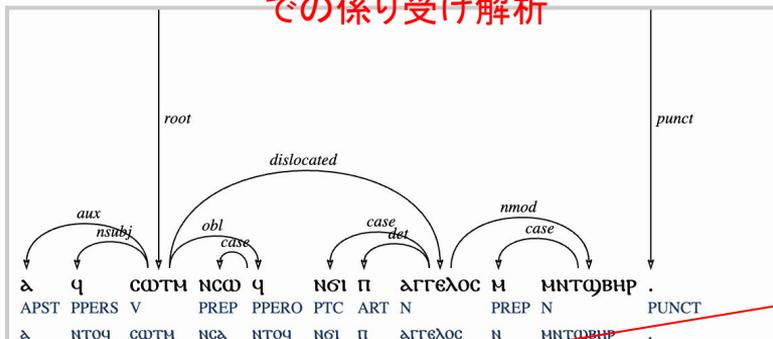
Result: XML / SGML での形態素分析

ⲁ<hi rend="red">ⲕ</hi>ⲥⲟⲩⲧⲙ ⲛⲥⲟⲩ ⲕⲓ  
 ⲓⲛⲁⲓⲛⲧⲉⲗⲟⲥ ⲓⲛⲓⲛⲧⲟⲩⲃⲏⲣ .  
 ⲟⲩ ⲧⲉⲧⲏⲣⲟⲩⲉ ⲛⲧⲟⲩⲛ ⲛⲉⲧⲕⲁⲙⲁⲛⲁⲥ  
 ⲉⲧⲣⲉⲕⲉⲓ ⲉⲣⲟⲩⲛ ⲛⲧⲉⲕⲁⲗⲓⲛ ⲛⲓⲛⲙⲁ  
 ⲉⲧⲟⲩⲁⲁⲃ ⲓⲛⲓⲛⲟⲩⲧⲉ ⲁⲗⲟⲩ  
 ⲛⲧⲉⲧⲏⲥⲟⲩⲟⲩ ⲓⲛⲉⲕⲣⲓⲣⲉ  
 ⲉⲧⲉⲛⲧⲟⲩⲛ ⲛⲉ ⲓⲛⲓⲛⲓⲛⲙⲟⲩⲛⲧⲏ  
 ⲓⲛⲉⲧⲏⲣⲓⲛⲉⲩⲉⲩⲉ  
 ⲛⲧⲁⲓⲛⲟⲩⲁⲥⲓⲥ ⲓⲛⲉⲓⲛⲟⲩⲟⲥⲟⲥ  
 ⲕⲉⲓⲉⲧⲏⲁⲥⲟⲩⲟⲩ ⲓⲛⲉⲓⲛⲟⲩⲧⲉ  
 ⲓⲛⲓⲛⲟⲩⲧⲉ .

コプト語テキスト入力

Syntax:

### Universal Dependencies での係り受け解析

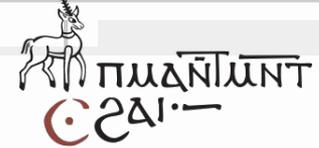


```

<norm_group orig_group="ⲁⲕⲥⲟⲩⲧⲙ" norm_group="ⲁⲕⲥⲟⲩⲧⲙ">
<norm xml:id="u1" pos="APST" lemma="ⲁ" func="aux" head="#u3" new_sent="true"
norm="ⲁ">
ⲁ
</norm>
<norm xml:id="u2" pos="PPERS" lemma="ⲛⲧⲟⲩⲕ" func="nsubj" head="#u3" orig="ⲕ" n
<hi rend="red">
ⲕ
</hi>
</norm>
<multiword multiword="ⲥⲟⲩⲧⲙ ⲛⲥⲟⲩ">
<norm xml:id="u3" pos="V" lemma="ⲥⲟⲩⲧⲙ" func="root" orig="ⲥⲟⲩⲧⲙ" norm="ⲥⲟⲩⲧⲙ">
ⲥⲟⲩⲧⲙ
</norm>
</norm_group>
    
```

Entities:

### エンティティ認識



ⲁ ⲕⲥⲟⲩⲧⲙ ⲛⲥⲟⲩ ⲕⲓ ⲛⲉⲧⲕⲁⲙⲁⲛⲁⲥ ⲓⲛⲓⲛⲧⲉⲗⲟⲥ ⲓⲛⲓⲛⲧⲟⲩⲃⲏⲣ . ⲟⲩ ⲧⲉⲧⲏⲣⲟⲩⲉ ⲛⲧⲟⲩⲛ  
 ⲛⲉⲧⲕⲁⲙⲁⲛⲁⲥ ⲓⲛⲉⲧⲕⲁⲙⲁⲛⲁⲥ ⲉⲧⲣⲉⲕⲉⲓ ⲉⲣⲟⲩⲛ ⲛⲧⲉⲕⲁⲗⲓⲛ ⲛⲓⲛⲙⲁ ⲉⲧⲟⲩⲁⲁⲃ ⲓⲛⲓⲛⲟⲩⲧⲉ  
 ⲧⲉⲧⲏⲥⲟⲩⲟⲩ ⲓⲛⲉⲕⲣⲓⲣⲉ ⲉⲧⲉ ⲛⲧⲟⲩⲛ ⲛⲉ ⲓⲛⲓⲛⲓⲛⲙⲟⲩⲛⲧⲏ ⲓⲛⲉⲧⲏⲣⲓⲛⲉⲩⲉⲩⲉ ⲛⲉ  
 ⲧⲁⲓⲛⲟⲩⲁⲥⲓⲥ ⲓⲛⲉⲓⲛⲟⲩⲟⲥⲟⲥ ⲕⲉ ⲓⲛⲉⲧⲏⲁⲥⲟⲩⲟⲩ ⲓⲛⲉⲓⲛⲟⲩⲧⲉ ⲓⲛⲓⲛⲟⲩⲧⲉ .



[Coptic NLP Service](#)

# ドイツのDHの特徴

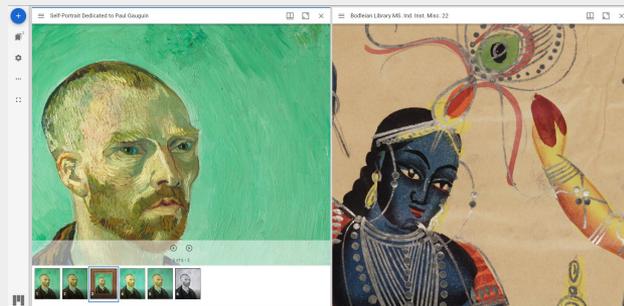
- 公的資金が多数投入されている
- 有期のプロジェクトで多数の学生とポスドク、PD、Juniorprofessur、Apl-Prof.が雇われている
  - 多くのプロジェクトは2-5年で終了するため、プロジェクトの終了とともに新しいプロジェクトに参加し、引越して「渡り歩く」キャリアの研究者が多い
  - エジプト学コプト学講座のように小さな専修では、無期雇用の研究者は専修のトップの教授のみで、その教授の指揮の元、複数の有期・無期雇用の教員・研究者が働いている
  - 学生はWissenschaftliche\*r Mitarbeiter\*in以外にも、Wissenschaftliche Hilftskraft、Studentische Hilftskraftなどが雇われる
- ITエンジニアをWMなどで雇うか、コンピュータ科学者と共同研究という形で組んでいるDHプロジェクトが多い
- 人文学のプロジェクトは、何らかの形でDHをしていることが多い
  - DHプロジェクトへのデータ登録などが学生を雇える仕事になっている

# ドイツのDHでの標準形式: TEIとIIIFとLODとCC

- TEI (Text Encoding Initiative)
  - 人文学のテキスト(写本・碑文・書簡・歴史的出版物の翻刻、校勘情報、言語学的コーパスなど)の構造化による機械可読化
  - TEI Guidelinesが定めたXML形式
  - 豊富な視覚化と分析ツール
  - 文献学系の多くのDHプロジェクトが使用
- IIIF (International Image Interoperability Framework)
  - 高精細画像を低負荷で表示すると同時に相互運用可能にする
  - メディアの比較とアノテーション
  - 最近では、画像だけでなく、音声・動画・3Dに対応
- LOD (Linked Open Data)
  - 柔軟な知識・メタデータ・データの蓄積・検索、セマンティック・ウェブ
- CC (Creative Commons)
  - ライセンスの標準形式: CC BY (二次利用の際に著作者の表示義務)など

# IIIF（国際的な画像の相互運用のための枠組み）

- 世界各地の有力な研究図書館により、様々な技術を結集して始められた**高精細画像相互運用技術**
- 大英図書館、オックスフォード大学ボドリアン図書館、フランス国立図書館 (Gallica)等、世界中の多くの著名な MLAK がデジタルアーカイブに採用
- 日本でも現在、様々な図書館、大学図書館、博物館、美術館などが IIIF を採用⇒急速に普及中
- **デジタルアーカイブ横断検索ポータル**で活用
  - ヨーロッパの Europeana
  - 日本のジャパンサーチ、Cultural Japan
- 面白い活用例: [Cultural Japan Self Museum](#)



## 国際的画像相互運用可枠組み・トリプルアイエフ International Image Interoperability Framework

学術的注釈

学術的比較

SAT Taishōzō Image DB [input a search word] All Search tags Previous search result

Vol.1-Vol.2-Vol.3-Vol.4-Vol.5-Vol.6-Vol.7-Vol.8-Vol.9-Vol.10-Vol.11-Vol.12 | 別紙 - | Jump to the page

大正新編大藏經圖像部第12b02巻

Annotations

羅刹 胎藏曼荼羅(子島寺藏本)  
(曼荼羅) 目の数: 二目 面相: 慈悲相: 閉口, 髻: 無, 鬘: 無  
胎面: 無, 額: 白毫無 持物: 棒(棒), 臂数: 二臂 坐法: 安坐: 右脚上, 履物: 無(無足) 袈身具: 耳飾: 無, 胸飾: 環珞, 臂釧: 環, 腕釧: 環, 足釧: 無, 台座: 蓮華, 光背: 環光: 無, 身光: 無, 月輪: 無, 臉の向き: 正面 タグ付け担当: 花澤明優美

月光菩薩 胎藏曼荼羅(子島寺藏本)  
(曼荼羅) 髮部: 結髮: 髻数

釋迦如來 胎藏曼荼羅(子島寺藏本) (曼荼羅) 目の数: 二目 面相: 慈悲相: 閉口, 胎面: 無, 額: 白毫有 臂数: 二臂 坐法: 結跏趺坐: 右脚上, 台座: 蓮華, 光背: 環光: 円光, 身光: 無, 炎光, 寶鬘, 額の向き: 正面 タグ付け担当: 津田勝英

東門 胎藏曼荼羅(子島寺藏本) (その他) 髮部: 頂上髪: 無, 目の数: その他 面相: 慈悲相: 閉口, 髻: 無, 額: 無, 胎面: 無, 額: 白毫有 持物: 無, 宝珠, 獅子, 鳥数: その他(正体が特定できないもの), 環珞, 臂数: その他 坐法: 結跏趺坐: 右脚上, 袈身具: 耳飾: 無, 胸飾: 無, 腕釧: 無, 足釧: 無, 台座: 蓮華, 光背: 環光: 円光, 身光: 有, 炎光, 額の向き: 正面 タグ付け担当: 黒崎夏央

羅刹 髮部: 頂上髪: 無, 目の数: 二目 面相: 慈悲相: 閉口, 髻: 無, 鬘: 無, 胎面: 無, 額: 無, 身坐法: 立: 両足立, 履物: 無(無足), 台座: 蓮華, 光背: 環光: 無, 身光: 無, 月輪: 無 胎面: 無 胎面: 無

THE SAT DAIZŌKYŌ TEXT DATABASE  
[https://21dzk.l.u-tokyo.ac.jp/SAT/index\\_en.html](https://21dzk.l.u-tokyo.ac.jp/SAT/index_en.html)

環, 足釧: 無, 台座: 蓮華, 光背: 環光: 円光, 身光: 有, 月

SIDEREVS  
NUNCIVS  
MAGNA, LONGEQVE ADMIRABILIA  
Spectacula pandens, fuscipendique proponens  
vnicuique, praeferrim vtro  
PHILOSOPHIS, atq; ASTRONOMIS, que 2  
GALILEO GALILEO  
PATRITIO FLORENTINO  
Patavini Gymnasij Publico Matematico  
PERSPICILLI  
Quorū 2  
Circū 10 VI S Stellarū dispersarū intervallicarū, atque periodicarū, celestiatarū mirabilium circumscripturarū; quos, nemini in hac vltiquē  
dicim cognoviss, nositissime Aulhor depræ-  
bendisse vtriusq; avicæ.

OVATOR PLANETIS  
MEDICEA SIDERA  
NVNCVPANDOS DECREVIT.

VENETI

gationem. M D C X.  
C. Trinitate.

+ (ここがメインですが) 相互運用



# オープンサイエンスの動向

- 論文を公開するだけでなく論文のデータも公開する
- 研究結果の検証性・再現可能性の確保
- 引用の効率性、インパクトファクタの測定可能性
- 研究論文、データにDOI（デジタルオブジェクト識別子）
- 各種リポジトリの発展・義務化
  - [Zenodo](#) (欧州連合CERN、OpenAIREなど)
  - [PubMed](#)
  - その他さまざまなりポジトリ
- 論文・データセットのオープンアクセス(OA)
- ライセンスの標準化・簡略化
  - [Creative Commons](#) ライセンス (CC BY 4.0、CC0など)
- 人文学は他分野と比べ遅れている(と言われる)



10.1021/ja803370x

# Canonical Text Service(CTS)

「Canonical Text Services (CTS) は、正準参照で引用されたテキストのパッセージを識別し、取得するためのプロトコルである。Canonical (旧称:Classical) Text Services仕様は、古典学やその他の文学分野で伝統的な「作品」と「引用」の概念を用いて、テキストの識別やテキストの断片を取得するためのネットワークサービスを定義している。」([Digital Classist Wiki](#)の和訳)

DOIは論文や研究で用いられるデータセットのための識別子だが、CTSは、**古典作品のパッセージ（章や節）のための識別子**を提供

**arabic\_translation:** Philippe Zaher

**collection:** Borgia Collection

**corpus:** shenoute.abraham

**document\_cts\_urn:**

urn:cts:copticLit:shenoute.abraham.monbxi:23-24

**entities:** gold

**identities:** gold

**language:** Sahidic Coptic

**msContents\_title\_n:** NA

**msItem\_title:** Abraham Our Father

**next:**

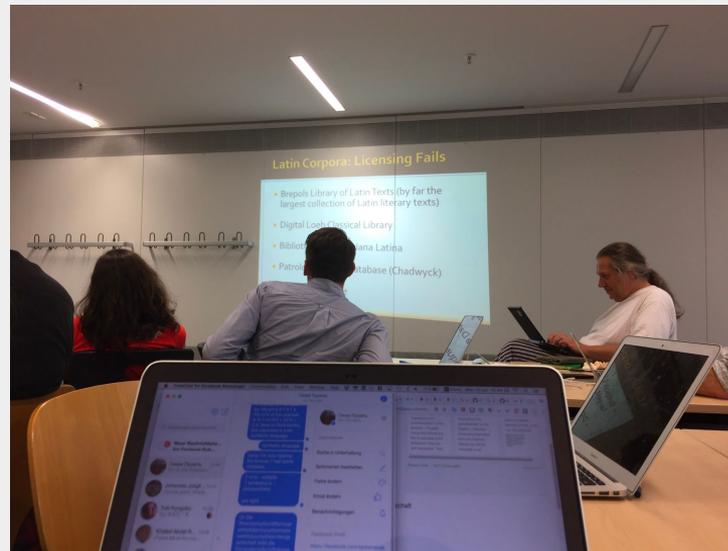
urn:cts:copticLit:shenoute.abraham.monbya:21-27

**origDate:** between 900 and 1200 C.E.

**origDate\_notBefore:** 0900

# Open Philology Project (Baumgardt et al. 2014)

- **ライプチヒ大学**を中心に2013-2018年に活動したプロジェクト(PI: Gregory Crane教授)
- OCR・HTR(手書きテキスト認識)、CTS、ラテン語・ギリシア語のためのTEIエディタ、ラテン語・ギリシア語のための係り受け解析エディタ、引用・引喩の機械探知、リンクト・オープン・データ、デジタル・エディションのプラットフォームなど様々な**ツールの開発を促進**し、数多くの**研究者・学生向けのワークショップ**を開発した(発表者はそのうちのほとんどのに参加)



Open Philologyコーパス開発のワークショップの風景 (2017年7月10日)

# 最後に：欧州のDigital Humanitiesのプロジェクトでよく言及されるFAIR原則

[NBDC 研究チーム 2020「データ共有の基準としてのFAIR原則」](#)

**F**indable

発見/検索可能

**A**ccessible

アクセス可能

**I**nteroperable

相互運用可能

**R**eusable

再利用可能

何人にも

フ

エ

ア

な研究環境

相互に繋がるデジタル技術を通して、

研究を研究者だけのものとせず、

一般の方でも資料の閲覧・研究文献の

引用をしながら研究成果を発表できる

オープンな人文学へ

デジタル研究基盤の  
整備が必要

# デジタルインフラストラクチャの充実

## DHを行うための様々なデジタルインフラストラクチャの充実

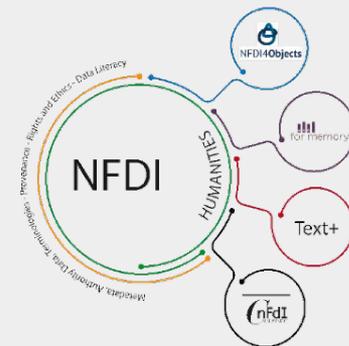
- デジタルインフラストラクチャプロジェクトで作られたツールなど
  - DARIAH-DE, DARIAH-EUなど (DARIAH: Digital Research Infrastructure for the Arts and Humanities)
    - [TextGrid](#)
    - [dariahTEACH](#)
  - CERN・OpenAIRE
    - [Zenodo](#)
  - CLARIN-Dなど (CLARIN: Common Language Resources and Technology Infrastructure)
    - [WebLicht](#) (CLARIN-Dインフラストラクチャ、チュービンゲン大学)
    - [WebAnno](#), INCEpTION (CLARIN-Dインフラストラクチャ、チュービンゲン大学)
- 通常の研究プロジェクトの助成で作られたツール
  - DFG
    - [ANNIS](#)
    - [CATMA](#) など多数

# DARIAH-DE, CLARIAH, そしてNFDI

ドイツにおける主要なデジタルインフラストラクチャプロジェクト



<https://de.dariah.eu/en/home>



# NFDI: Nationale Forschungsdaten Infrastruktur

## RUNDE 1 (2020/10-)



### NFDI4Culture

Konsortium für Forschungsdaten zu materiellen und immateriellen Kulturgütern

[Mehr erfahren](#)

## RUNDE 2 (2021/10-)



### Text+

Sprach- und textbasierte Forschungsdateninfrastruktur

[Mehr erfahren](#)



Gemeinsame  
Wissenschaftskonferenz  
GWK



GEFÖRDEBT VOM

Bundesministerium  
für Bildung  
und Forschung

## RUNDE 3 (2023/5-)

- **NFDI4Memory**: Konsortium für historisch arbeitende Geisteswissenschaften
- **NFDI4Objects**: Forschungsdateninfrastruktur für die materiellen Hinterlassenschaften der Menschheitsgeschichte
- **NFDI4CS**: Nationale Forschungsdateninfrastruktur für und mit Computer Science

# 資料のデジタル化・デジタル人文学の参考書のご紹介

より効果的にテキストデータを蓄積・共有し、広く世界につなげる  
これからのテキストの作り方・広げ方!

## 人文学のための テキストデータ 構築入門

TEIガイドラインに準拠した取り組みにむけて

【監修】  
伊藤智文  
宮内一夫  
長岡一雄  
小川 誠  
片岡博子  
森 雅彦  
小島明雄  
宮内昭博  
長岡一夫  
中村 文  
藤 野一  
丸島正典

【発行】一般社団法人 人文情報学研究所  
【著者】石田友梨 / 大向一輝 / 小島綾乃  
永崎研彦 / 宮川 創 / 渡邊要一郎

Contents  
第1部 総論  
第2部 実践編  
第3部 事例編：テキストデータ構築の最新事情1  
第4部 事例編：テキストデータ構築の最新事情2  
TEI活用事例紹介

私たちの読みを残し、共有し、たどれるように。このことで、どう読んだかを次世代に継承する――

『人文学のためのテキストデータ構築入門：TEIガイドラインに準拠した取り組みにむけて』

文学通信：2022年、  
3300円



『欧米圏デジタル・ヒューマニティーズの基礎知識』

文学通信：2021  
年、3080円

ウェブマガジン『人文情報学月報』で「欧米・中東デジタル・ヒューマニティーズ動向」連載中!

# ドイツにおけるDHのまとめ

- プロジェクトの助成および**デジタルインフラストラクチャ**が充実している
- **オープンサイエンス** (論文と研究データを公開) が広く進んでいる
- 人文学の分野でも、**オープンアクセス**の論文が増えている
- 文献資料を公開する**デジタルアーカイブ**および**検索プラットフォーム**が増え  
てきている
- 消滅の危機にある言語の分野でもデジタルアーカイブ化が進む
- 論文や研究データの識別子**DOI**
- 画像・音声・動画の国際的相互運用枠組みの**IIIF**
- テキストデータの構造化の世界標準**TEI**
- データのやりとりをオープンにし、柔軟な検索を実現する**LOD (RDF)**
- 人工知能 (**ChatGPT**など) の活用
- より**オープン**で**FAIR**な研究環境

今回述べることができたのは、ドイツのDHのごく一端

# ドイツにおけるDHに関する情報ブログ: [DHd-Blog](#)



[Login](#) [Registrierung](#) [Beiträge](#) [Kommentare](#) [DARIAH-DE](#) [Datenschutz](#) [Impressum](#)

[Blog](#) [Über uns](#) [Redaktion](#)

Suchen

## Einladung zum Text+ Plenary 2023: Connecting People and Data

Veröffentlicht von Lukas Weimer am 16. Juni 2023



Am 28. und 29. September 2023 findet an der Niedersächsischen Staats- und Universitätsbibliothek Göttingen das zweite Plenary des NFDI-Konsortiums Text+ statt, zu dem die Community mit Mitarbeitenden und Delegierten aus Fachverbänden und -verbänden zusammenkommt. Alle Interessierten sind herzlich eingeladen! Die Tagung steht unter dem Motto „Connecting People and Data“. Als Konsortium der Nationalen Forschungsdateninfrastruktur (NFDI) hat Text+ zum [...]



Annotation **BBAW Berlin**  
Call for Papers **CLARIAH-DE**  
CLARIN-D Community  
**DARIAH-DE** DARIAH-EU DH-  
Kolloquium **DH-Videos**  
DHD DHd-Kanal DHd2022  
**Digitale Edition**  
Digitale  
Geschichtswissenschaft  
**Digital Humanities**  
Digitalisierung  
**Forschungsdaten**  
Forschungsdateninfrastruktur  
**Forschungsdatenmanag**  
Forschungsinfrastruktur Geschichte  
Graz Göttingen IDE Konferenz  
Kunstgeschichte NFDI **Open**

# ありがとうございました！ Vielen Dank!

<https://somyagawa.com/> (ホームページ1)

<https://somyagawa.de/>

(ホームページ2)

[miyagawa.so.kb@u.tsukuba.ac.jp](mailto:miyagawa.so.kb@u.tsukuba.ac.jp) (Eメール)

