

第139回 人文科学とコンピュータ研究会

年譜資料の構造化と可視化 —森鷗外「抽斎年譜」を例に—

鈴置 恵子（東京大学情報システム部）

市太 佐知（人間文化研究機構）

西口 真梨奈（東京都立中央図書館）

福嶋 彩乃（東京大学経済学図書館）

星合 耀子（国文学研究資料館）

村田 祐菜（国立国会図書館）

吉村 雄太（国立国会図書館）

はじめに

- ・東京大学総合図書館所蔵『抽斎年譜』を対象にTEI/XML形式でテキストを構造化し、可視化サイトを構築した事例を報告する
- ・本発表の取組みは、複数機関の図書館職員による自主勉強会*の成果

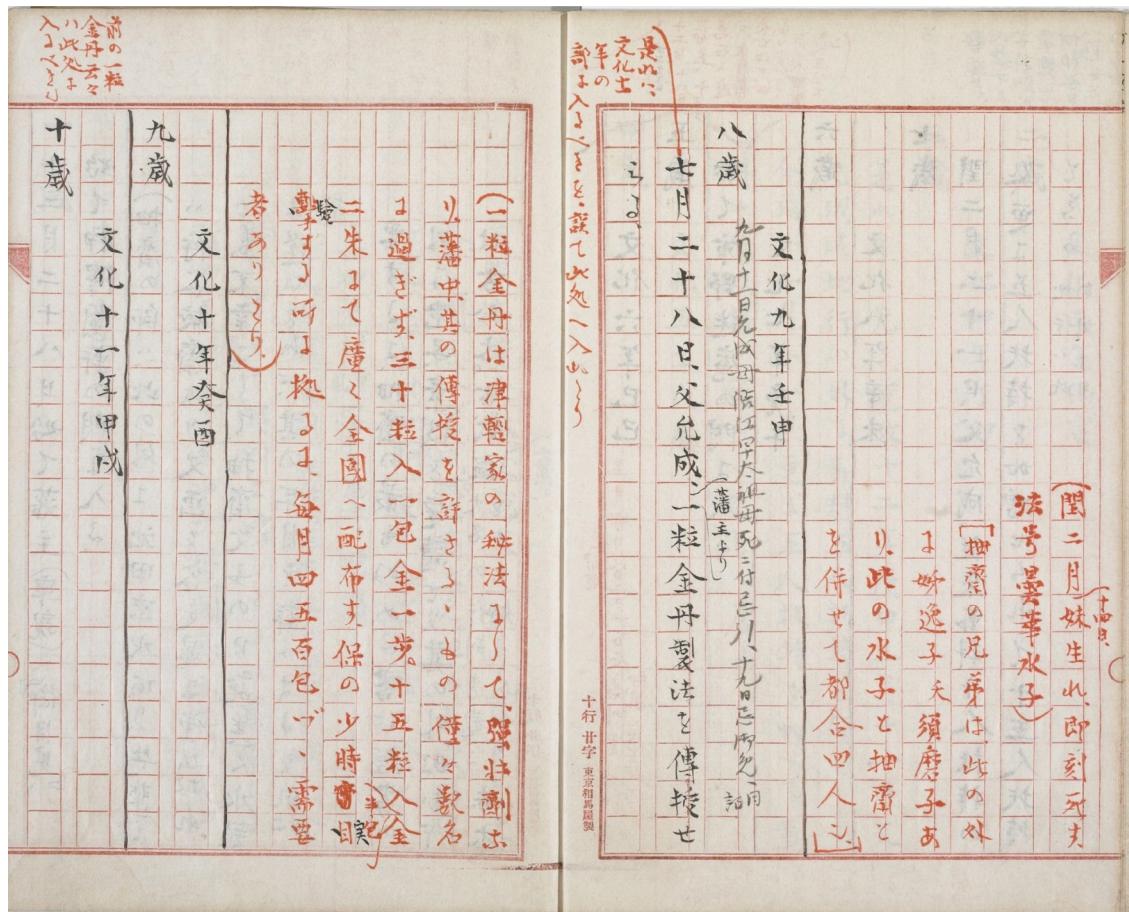
*若手図書館員DH勉強会 <https://dhlibrarianstudygroup.github.io/website/docs/about/>



発表内容：

- ・森鷗外による書入れ本『抽斎年譜』の概要
- ・構造化の方針・共同編集作業の方法
- ・構造化の対象とする要素
- ・TEIマークアップ方法
- ・可視化サイトの構築

書入れ本『抽斎年譜』の概要



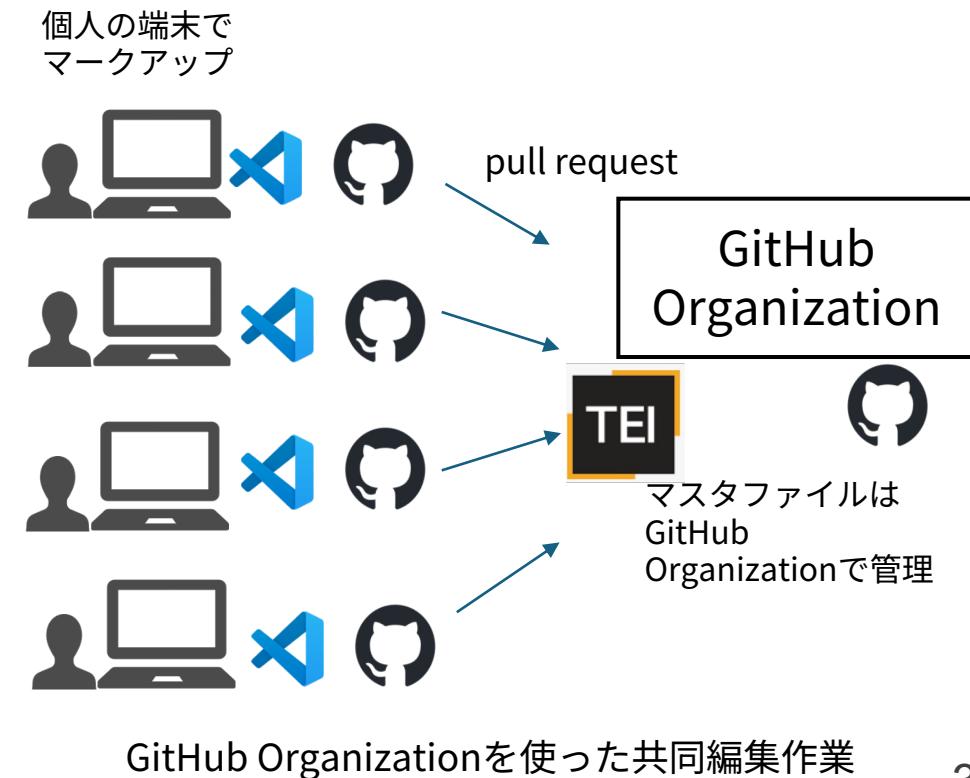
- 森鷗外が史伝『渋江抽斎』を執筆する際に抽斎の嗣子・渋江保に制作を依頼した年譜資料
- 渋江抽斎(1805-1858)：
弘前藩の侍医・考証家・書誌学者
- 史伝小説である『渋江抽斎』は、本資料に大きく依拠するとされる
→『抽斎年譜』と『渋江抽斎』のテキストを対照し、年譜を典拠として執筆された箇所を可視化したい
- 制作者である渋江保自身が書き込みを加えている。また、鷗外自筆の書入れも見られる。
→注の位置や朱筆/墨筆の別、書入れを行った人物の情報を記述したい

東京大学総合図書館所蔵『抽斎年譜』（画像の公開元：鷗外文庫書入本画像データベース）

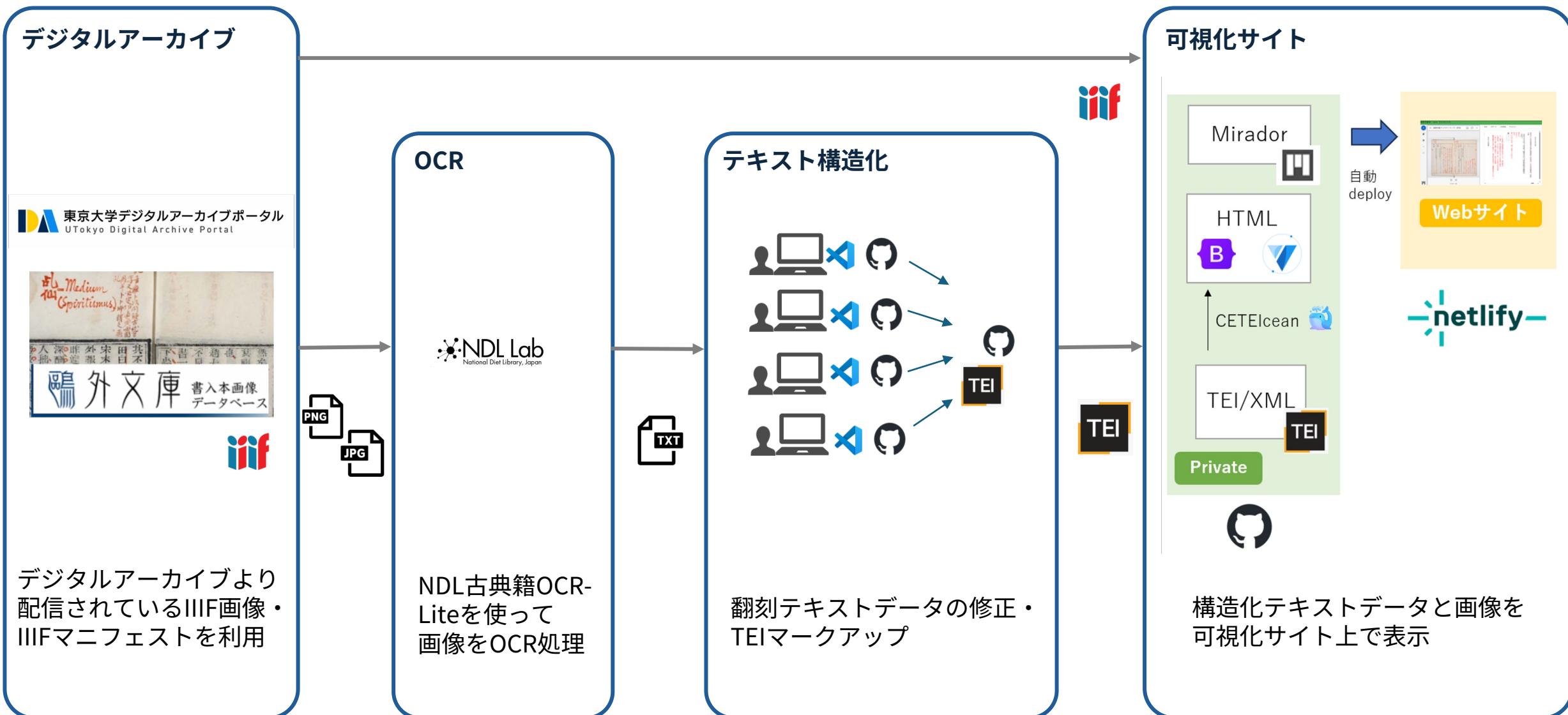
<https://iiif.dl.itc.u-tokyo.ac.jp/repo/s/ogai/document/60d193c1-d814-42c3-90c0-a5caf5975e1a#?c=0&m=0&s=0&cv=0&xywh=-532%2C555%2C4279%2C2552>

テキスト構造化の作業方針

- TEI P5ガイドラインに準拠してテキストを構造化する
- 割り注やルビは、TEI協会東アジア/日本語分科会の日本語古典籍TEI本文データ作成要領を参照してマークアップする
- 可能な限り、オープンソースのツールを用いる
 - エディタ
Oxygen XML Editorは価格面のハードルがあるため、Visual Studio CodeでTEI/XMLの編集ができるよう拡張機能を導入
 - バージョン管理・共同編集
GitHub OrganizationリポジトリにTEI/XMLを置き、個人アカウントから更新
 - OCR
NDL古典籍OCR-Liteを使って画像からテキストデータを取得→人手で修正



可視化サイト構築の概観



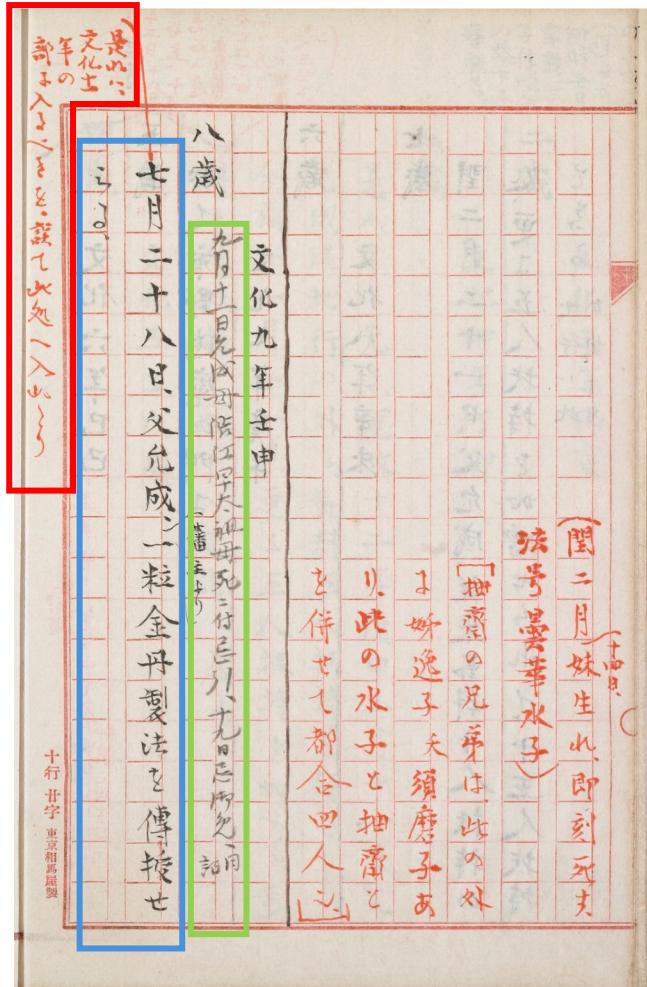
<note style="color:red" place="top">
 <div>

 <head style="#m3"><date when="1855" when-custom="安政2年">安政二年乙卯</date></head>
 <head>五十一歳</head>
 <head>
 • style="#m3": インデントの升目の数（ここでは3マス行頭が下がっている）
 • 日付のマークアップ:
 <date when="1855" when-custom="安政2年">安政二年乙卯</date>
 </head>
 <p style="#m2">
 此頃、建護して藩士一同甲冑と所有せしも
 もの悉く國勝手に落さへり。又藩侯一家並み藩士江戸定府の
 扶持へ終身の扶持より被仰付
 </p>
 <head>五十三歳</head>
 <head>五十二歳</head>
 <head>安政三年丙辰</head>
 <p style="#m5">
 此の年大地震。従来抽籠一切
 </p>
 <head>安政四年丁巳</head>
 <note style="color:red;#m10" rend="割注">
 十二月二十八日、< milestone unit="wbr"/>嫡子願濟
 </note>
 <note style="color:red;#m10" place="bottom">
 割書：
 タグは<note>、属性はrend="割注"とする。改行位置に< milestone unit="wbr"/>を入れる。
 </note>
 <div>
 <head>三月十九日翠馨生る。慶</head>
 <note style="color:red;#m10" place="bottom">
 元年乙卯六月二十日
 </note>
 </div>
 <div>
 <head>三月十九日翠馨生る。慶</head>
 <note style="color:red;#m10" place="bottom">
 元年乙卯六月二十日
 </note>
 </div>

年譜の構造、形態的な要素（見た目）

- <div> : 年ごとのかたまり
- <head> : 年、年齢
- <p> : 本文
- @style : インデント
- <note> : 注釈
- @style : 朱筆
- @place : 位置
- < milestone />、@rend : 割注
- <lb/> : 改行

TEIマークアップ—注・書入れ



本文

欄外注

xml:idで対応を示す

<lb/><p style="#m1" xml:id="cn05_01">七月二十
八日、父允成、藩主より一粒金丹製法を傳授せ
<lb/>らる、</p>

<note place="margin" style="color:red" corresp="#cn05_01">
是れハ、
<lb/>文化十一
<lb/>年の
<lb/>部に入るべきを、誤て此處へ入れたり</note>

style属性で朱筆を示す

鷗外による書入れであることを示す

<note hand="#ougai">九月十一日允成母渋江早太祖母死ニ付忌引、十九日忌御免、日記</note>

TEIマークアップ－インデント

<tagsDecl>

<teiHeader>配下、<encodingDesc>中の
<tagsDecl>にインデント情報をまとめて記述しておく

```
<encodingDesc>      インデントの種類ごとにIDを付与する
  <tagsDecl>
    <rendition xml:id="m1">margin-top: 1em;</rendition>
    <rendition xml:id="m2">margin-top: 2em;</rendition>
    <rendition xml:id="m3">margin-top: 3em;</rendition>
    <rendition xml:id="m9">margin-top: 9em;</rendition>
    <rendition xml:id="m10">margin-top: 10em;</rendition>
    <rendition xml:id="m11">margin-top: 11em;</rendition>
  </tagsDecl>
</encodingDesc>      三文字分の空白を示す
```

<text> (本文)

style属性でxml:idを記述して左記のインデント情報を参照する

xml:idで参照

```
<div>
  <head style="#m3">年代</head>
  <head>抽斎の年齢</head>
  <note style="color:red;#m10">原稿用紙枠内の朱筆注記</note>
  <p style="#m5" xml:id="note01_01">
    <lb/>本文1行目
    <lb/>本文2行目
    <lb/>本文3行目
  </p>
  <note style="color:red" place="注記の場所"
        corresp="note01_01">原稿用紙枠外の朱筆注記</note>
</div>
```

可視化サイト

抽斎年譜TEI Home 青空文庫との比較

三 抽斎年譜(チュウサイ ネンブ). [写本]

本文 メタデータ 人物情報 タイムライン

抽斎又
越後
獅子角賞
兵衛を
好み、毎日
中の口炎
聞の數多に於
て嘗はし
め、在宅の
時八、(何等
用事あり
とも、必ら
ばそれを
打捨て置き、)
親から出
て、見
物したり、
十二月十六
日、藩主
任侍従
六月十六
日、藩主大広間
着坐を命
せらる

九歳

九月十一日允成母涉江早太祖母死一付忌引、十九日忌御免、日記
九月二十八日、父允成、「藩主より」一粒金丹製法を傳授せらるる、
八歳

文化九年壬申

(一粒金丹は津輕家の秘法にして、強壯剤なり、藩中、其の傳授を許さるゝもの僅々数名に過ぎず、三十粒入一包金一步。十五粒入半包)金
二朱にて廣く全國へ配布す、保の少時由來[実験]する所に據るに、毎月四五百包づゝ需要者ありたり、)

文化十年癸酉

九歳

九月二十八日、父允成、一粒金丹製法を傳授せり。
「抽斎の兄弟は、此の外に姉逸子夫須磨子あり、此の水子と抽斎とを併せて都合四人也。」

九歳

九月二十八日、父允成、「藩主より」一粒金丹製法を傳授せらるる、
八歳

文化九年壬申

(一粒金丹は津輕家の秘法にして、強壯剤なり、藩中、其の傳授を許さるゝもの僅々数名に過ぎず、三十粒入一包金一步。十五粒入半包)金
二朱にて廣く全國へ配布す、保の少時由來[実験]する所に據るに、毎月四五百包づゝ需要者ありたり、)

文化十年癸酉

九歳

九月二十八日、父允成、「藩主より」一粒金丹製法を傳授せらるる、
八歳

文化九年壬申

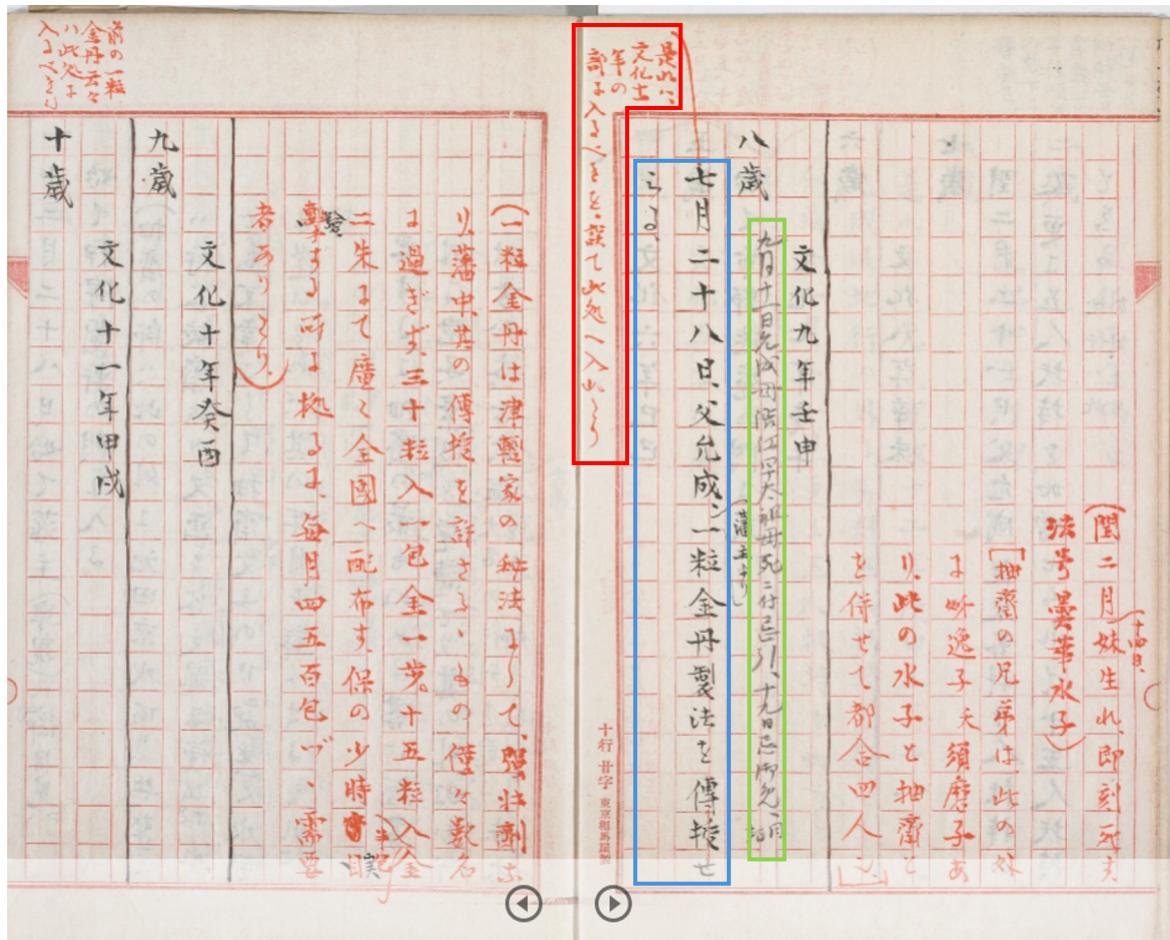
(一粒金丹は津輕家の秘法にして、強壯剤なり、藩中、其の傳授を許さるゝもの僅々数名に過ぎず、三十粒入一包金一步。十五粒入半包)金
二朱にて廣く全國へ配布す、保の少時由來[実験]する所に據るに、毎月四五百包づゝ需要者ありたり、)

文化十年癸酉

九歳

可視化サイト－注・書入れの表示

III F画像ビューワ

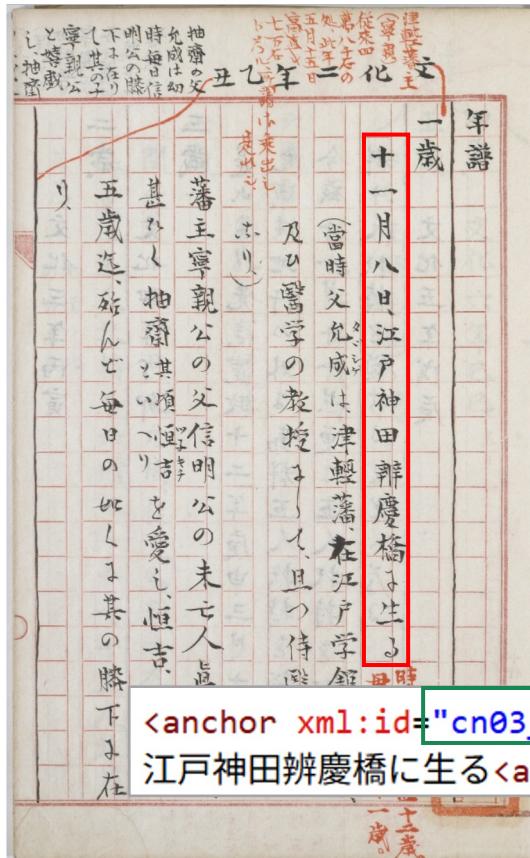


テキストビューワ

| 本文 | メタデータ | 人物情報 | タイムライン |
|--|-------|------|--|
| <p>是れ八 文化十一 年 金丹云々 ハ此處に 入るべきを 語て此處へ入れた り</p> <p>抽薪又 越後 獅子ヶ角宣 兵衛を 好み 每日 中の口炎 關の敷臺に於 て舞はし</p> <p>九月十一日允成母浜江早太祖母死ニ付忌引、十九日忌御免、日記</p> <p>（一粒金丹は津輕家の秘法にして、強壮剤なり、藩中、其の傳授を許さるゝもの僅々數名に過ぎず、三十粒入一包金一步。十五粒入〔半包〕金二朱にて廣く全國へ配布す、保の少時由</p> <p>擊実驗する所に拵るに、毎月四五百包づゝ需要者ありたり、）</p> | | | <p>九月十一日壬申 文化九年壬申</p> <p>九月十一日允成母浜江早太祖母死ニ付忌引、十九日忌御免、日記</p> <p>（一粒金丹は津輕家の秘法にして、強壮剤なり、藩中、其の傳授を許さるゝもの僅々數名に過ぎず、三十粒入一包金一步。十五粒入〔半包〕金二朱にて廣く全國へ配布す、保の少時由</p> <p>擊実驗する所に拵るに、毎月四五百包づゝ需要者ありたり、）</p> |
| <p>九歳</p> <p>文化十一年甲戌</p> <p>九歳</p> <p>文化十一年癸酉</p> <p>八歳</p> <p>文化九年壬申</p> <p>七月二十八日、父允成、一粒金丹製法を傳授せ</p> <p>閏二月、姫生山即刻元美 法号曼華水子</p> <p>「抽薪の兄弟は此の外 よ幼逸子天須磨子あり、此の水子と抽薪を 侍せて都合四人」</p> <p>九歳</p> <p>文化十年癸酉</p> | | | |

TEIマークアップ—『渋江抽斎』との対応箇所

『抽斎年譜』



<anchor xml:id="cn03_01_s"/>十一月八日、
江戸神田辨慶橋に生る<anchor xml:id="cn03_01_e"/>

森鷗外『渋江抽斎』

文青空
庫

その十二

抽斎は文化二年十一月八日に、神田弁慶橋に生れたと保さんがいう。これは母五百の話を記憶しているのである。父允成は四十二歳、母縫は三十一歳の時である。その生れた家はどの辺であるか。弁慶橋というのは橋の名ではなくて町名である。当時の江戸分間大絵図というものを閲するに、和泉橋と新橋との間の柳原通の少し南に寄って、西から東へ、お玉が池、松枝町、弁慶橋、元柳原町、佐久間町、四間町、大和町、豊島町という順序に、町名が注してある。そして和泉橋を南へ渡って、少し東へ偏って行く通が、東側は弁慶橋、西側は松枝町になっている。この通の東隣の筋は、東側が元柳原町、西側が弁慶橋になっている。わたくしが富士川游さんに借りた津軽家の医官の宿直日記によるに、允成は天明六年八月十九日に豊島町通横町鎌倉横町家主伊右衛門店を借りた。この鎌倉横町というのは、前いった図を見るに、元柳原町と佐久間町との間で、北の方河岸に寄った所にある。允成がこの店を借りたのは、その

<anchor corresp="#cn03_01_s"/>抽斎は文化二年十一月八日に、
神田弁慶橋に生れた<anchor corresp="#cn03_01_e"/>

xml:idで参照 (終了位置)

*対応箇所は『日本近代文学大系12 森鷗外集 II』(角川書店、1974年)の補注に拠る

TEIマークアップ—本文と対応するIIIF画像

<facsimile>

```
<facsimile sameAs="https://iiif.dl.itc.u-tokyo.ac.jp/repo/iiif/60d193c1-d814-42c3-90  
c0-a5caf5975e1a/manifest">  
  <surface xml:id="ut02" ulx="0" uly="0" lrx="4256" lry="3308" sameAs="https://iiif.  
dl.itc.u-tokyo.ac.jp/repo/iiif/60d193c1-d814-42c3-90c0-a5caf5975e1a/canvas/p3">  Canvas  
    <label>[3]</label>  
    <graphic mimeType="application/json" url="https://iiif.dl.itc.u-tokyo.ac.jp/repo/iiif-img/6981"/>  
    <zone lrx="4256" lry="3308" ulx="2128" uly="0" xml:id="zone_004"/> ←画像の右半分 (=1丁ウラ)  
    <zone lrx="2128" lry="3308" ulx="0" uly="0" xml:id="zone_005"/> ←画像の左半分 (=2丁オモテ)  
  </surface>
```

IIIF Manifest URI

Canvas

←画像の右半分 (=1丁ウラ)
←画像の左半分 (=2丁オモテ)

半丁単位のIDを付与

<text> (本文)

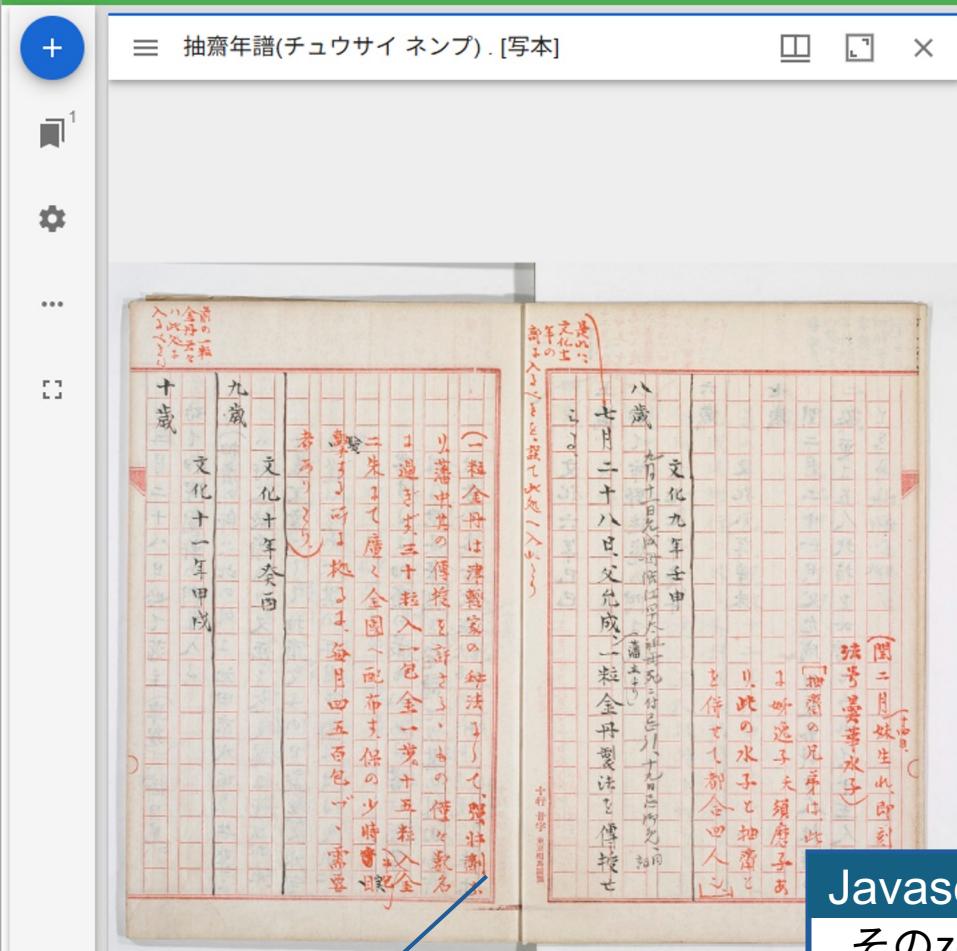
xml:idで参照

```
<text>  
  <body style="writing-mode:vertical-rl">  
    <pb n="3" corresp="#zone_005" facs="https://iiif.dl.itc.u-tokyo.ac.jp/repo/  
    iiif-img/6981/0,0,2128,3308/full/0/default.jpg"/>  
  <head>年譜</head>
```

IIIF Image API のURL記法で半丁分の領域を指定

可視化サイト —本文と対応するIIIF画像

抽斎年譜TEI Home 青空文庫との比較



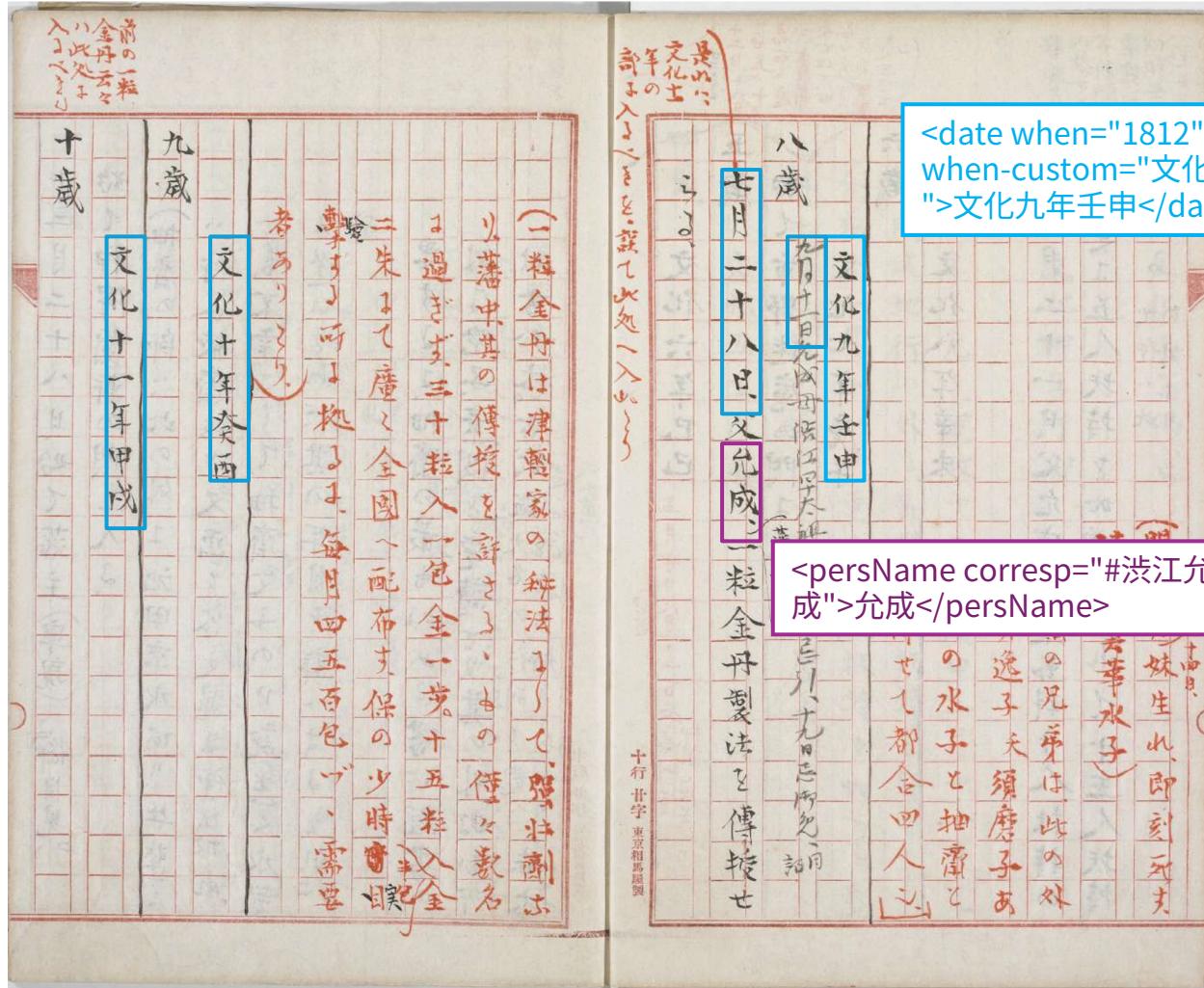
本文と対応する画像コマに遷移

Javascript 半丁ごとに挿入された<pb/>タグの@correspでzoneIDを参照、そのzoneIDの<zone>要素と同じ<surface>内の<graphic>に記述されているcanvasURIを取得してMiradorに渡す

本文テキストの半丁ごとに設置されたボタンを押下

〔閏二月十四日〕妹生れ
法号曼華水子

TEIマークアップ—時間情報・人物情報



時間と人物

<date when="1812" when-custom="文化9年壬申">文化九年壬申</date>

@when : 西暦

@when-custom : 和暦

<persName> : 人物

@corresp : 末尾の人物リストをxml:idで参照

xml:idで参照

```
<person xml:id="渋江允成">
  <persName>
    <surname>渋江</surname><forename>允成</forename>
  </persName>
  <note><p>渋江抽斎の父。</p></note>
</person>
```

人物情報の詳細は<back>の<listPerson>に記述

可視化サイト – 時間情報の表示

本文 メタデータ 人物情報 タイムライン

The screenshot shows a vertical timeline with colored dots indicating event points. A green dot is at the top, followed by grey dots. The first grey dot corresponds to the event on 1845-01-23. A blue arrow points from this event box to a red note box at the bottom.

- 1845年（弘化元年十二月十六日）
- 1845-01-23（弘化元年十二月十六日）
十二月十六日日、藩主叙四品、
- 1844-04-29（弘化元年三月十二日）
③三月十二日医学館講師たるべき台命ありたる旨、土井大炊頭より達せらる、
- 1844-06-14（弘化元年四月二十九日）
④四月二十九日、自今年始、八朔、五節句、月並御礼御城江戸城江戸城へ罷出べき旨、命ぜらる、

<note place="top" style="color:red"><date when="1845-01-23" when-custom="弘化元年十二月十六日">
十二月十六日</date><lb/>日、藩主<lb/>叙四品、</note>

『抽斎年譜』は編年体構成をとるが、出来事の記載は必ずしも発生順に並んでおらず、前後することがある
→出来事を西暦順にタイムライン表示したい

<date>の@whenに記述した西暦の情報をもとに、出来事の発生順にタイムライン表示

可視化サイト —人物情報の表示

本文 メタデータ 人物情報 タイムライン

渋江抽斎
生年：文化2年11月8日
没年：安政5年8月29日
VIAF：<https://viaf.org/viaf/18576420>

本文中の参照箇所を表示

参照 1 参照 2 参照 3 参照 4 参照 5 参照 6 参照 7 参照 8 参照 9 参照 10
参照 11 参照 12 参照 13 参照 14 参照 15 参照 16

森立之
VIAF：<https://viaf.org/viaf/44186331>

土井利位
VIAF：<http://viaf.org/viaf/251432551>

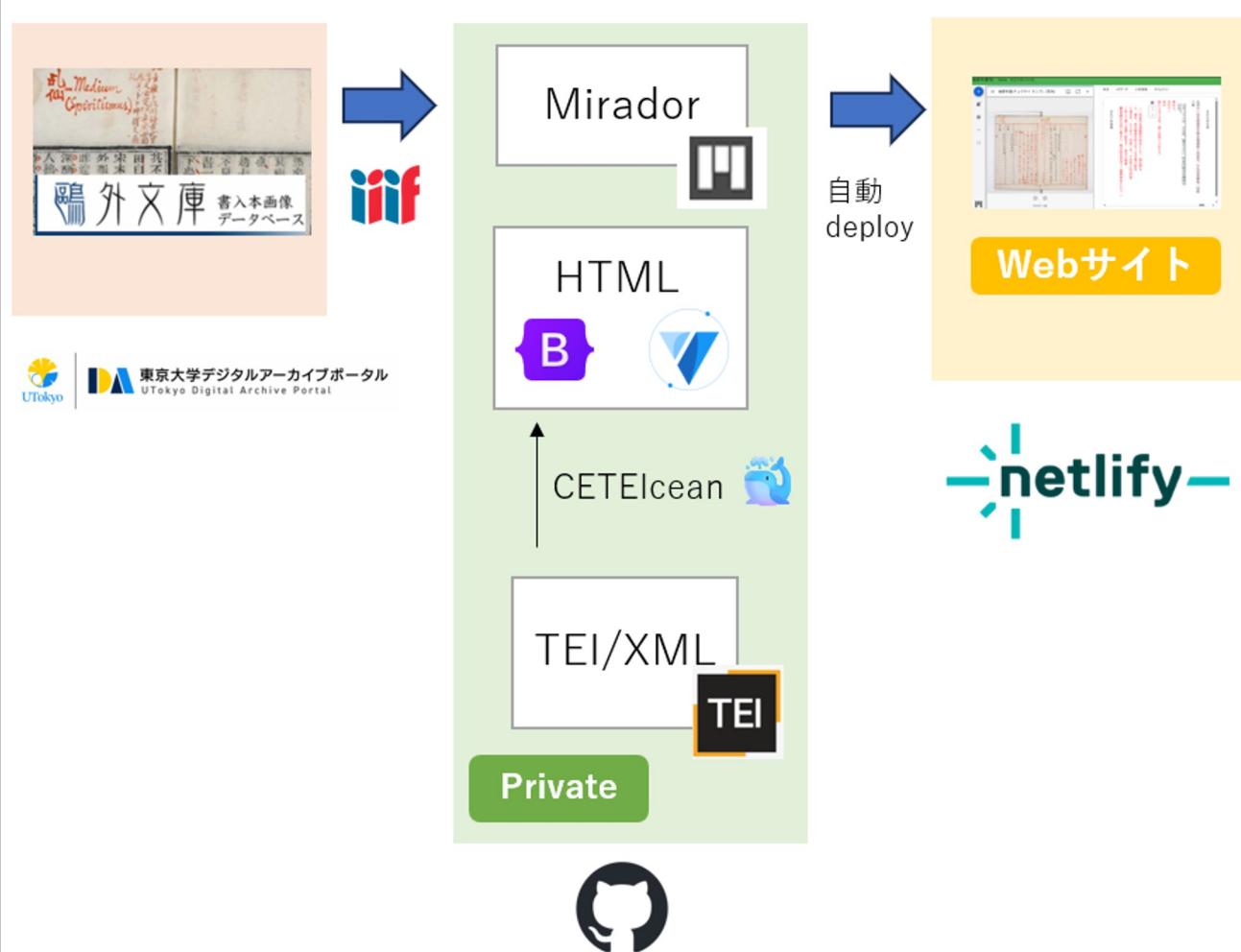
<listPerson>の人物情報を表示

『抽斎年譜』には、抽斎の血縁・友人・弘前藩の人物等が多数登場する。抽斎は四人の妻との間に七男七女をもうけており、年譜に登場する各人物が抽斎とどのような家族関係にあるのかわかりにくい

→人物情報を可視化したい（家系図やネットワークグラフでの実装を検討中）

```
<listPerson>
  <person>
    <persName xml:id="渋江抽斎">
      <surname>渋江</surname><forename>抽斎</forename></persName>
      <idno type="VIAF">https://viaf.org/viaf/18576420</idno>
      <birth>文化2年11月8日</birth>
      <death>安政5年8月29日</death>
    </person>
    <person xml:id="森立之">
      <persName>
        <surname>森</surname><forename>立之</forename>
      </persName>
      <idno type="VIAF">https://viaf.org/viaf/44186331</idno>
    </person>
  </listPerson>
```

可視化サイトの構成



- なるべく ミニマムな構成 に
 - HTML、CSS、Javascriptで静的サイトを構成
 - TEIやiiifの技術エコシステムを利活用
- CETElcean
 - JavaScriptライブラリ
 - TEI/XMLを読みこみ、HTML要素として扱える
- Mirador
 - iiifビューワーの一つ
 - 東大図書館で公開されている『抽斎年譜』iiifManifestのManifestURIを渡すと画像を表示
- GitHubリポジトリに資材を配置
→Netlifyで自動デプロイ

おわりに

- ・構造化・可視化の方法を共有する意義
 - ・諸本や原拠本を参照し、対応箇所を比較する過程は文献を扱う研究の基礎
 - ・エコシステムへの貢献：本発表の取組みでも先行事例を参考に機能を実装
- ・図書館所蔵資料の利活用促進：研究に資するデータを提供するために
 - ・どのような情報をどの深度までマークアップするか？
資料の性格や利用者のリサーチクエスチョンによって記述したい情報は異なる
 - ・内容に踏み込んだマークアップには専門知識とコストが必要
 - ・欧米/日本における図書館員と人文情報学との関わり方
- ・『抽斎年譜』のTEI/XMLと可視化サイトの公開に向けて

謝辞：TEIガイドラインに準拠したテキストの構造化については公益財団法人渋沢栄一記念財団情報資源センターに、可視化サイトの実装については東京大学史料編纂所の中村覚助教に、多大なご助言とご協力を賜りました。ここに謹んで感謝の意を表します。