

Good things come in small packages: designing distributed editions and tools for the age of FAIR data

Daniel Paul O'Donnell (Lethbridge)
Roberto Rosselli Del Turco (Torino)

Assemblée Générale du consortium Cahier
27 Nov. 2020

Part 1.

The problem of Humanities Data

Traditionally, humanists resist speaking of “data”

- Instead they tend to describe the raw material of their work as Primary and Secondary sources and “Readings”
 - “Primary sources” = Texts, artifacts, objects of study
 - “Secondary sources” = Works of other scholars (often based on “Primary sources”)
 - “Readings”
 - Passages, extracts, quotations for interpretation or support
 - Interpretation, the end product of research (literary study)

These definitions don't map easily onto "data"

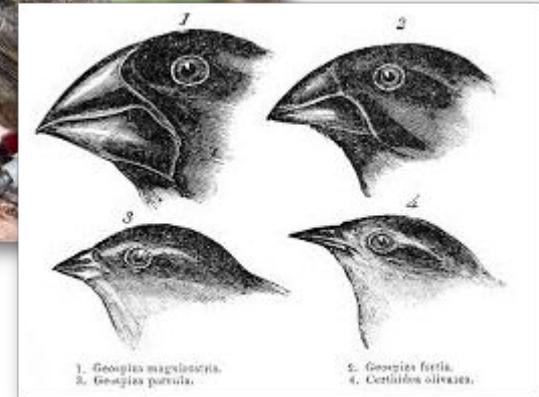
- "Primary sources"
 - Can be originals (i.e. the artifact itself)
 - Mediated/contextualised in some way: edition, transcription, photograph
 - Can be "Primary" in one context and "Secondary" in another
 - E.g. a classic work of philosophy that is both historically and philosophically important
 - Can be simultaneously "Primary" and "Secondary"
 - Representation of an object (primary) can be itself a work of scholarship (secondary)
 - Are debatable
 - Picture vs. Frame, Book vs. Binding
 - Often Humanities is about whether something *is* "data"

These definitions don't map easily onto "data"

- Likewise "readings" (in the sense of extracts from primary sources) are both "observations" and "representations"
 - You choose "readings" to illustrate your argument, but these are examples rather than notes
 - You pick the examples that support the argument rather than record what happens after an intervention
 - We may make notes that we don't use
 - But these are not the same as notes made as the result of a system, method based observation
 - No tradition of (or reason for) saving notebooks: integrity doesn't depend on reproducibility

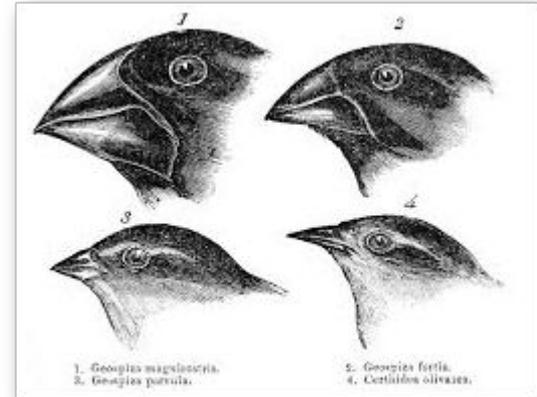
Darwin as an example... What was his data?

- To see this, let's think about what Darwin's data was when he went to the Galapagos Islands
 - Was it the finches?
 - Was it his notes and observations about the finches?



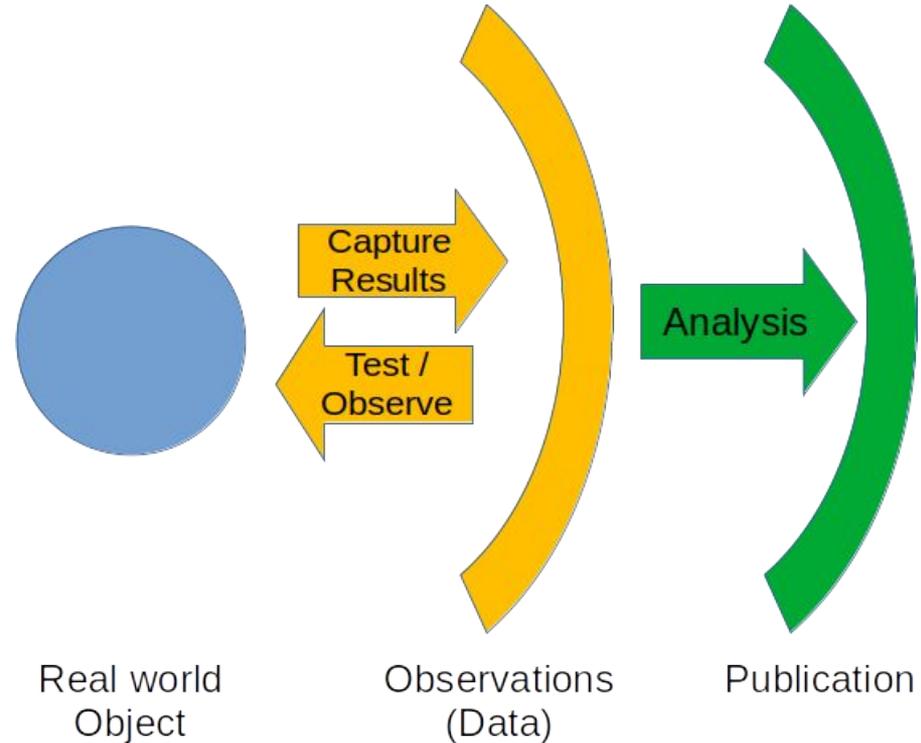
Darwin as an example... What was his data?

- In fact, it was his notes — the record of his observations
- “Data” = “represent[ation of] information in a formalized manner suitable for communication, interpretation, or processing” (NASA 2012);
“the facts, numbers, letters, and symbols that describe an object, idea, condition, situation, or other factors” (NRC 1999)



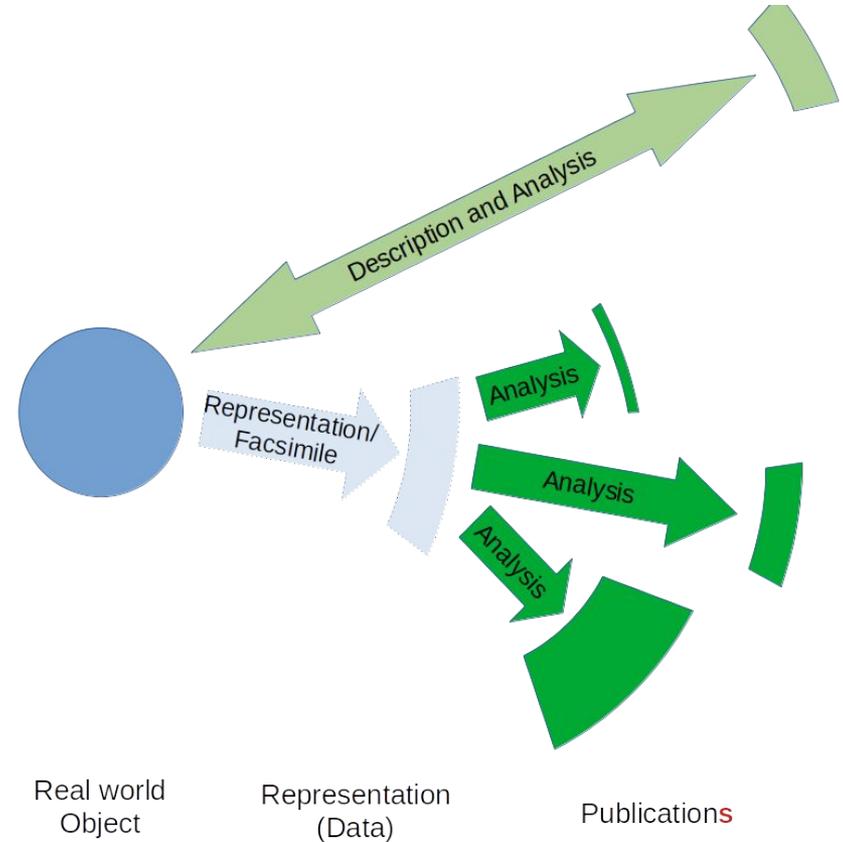
Darwin as an example... What was his data?

- What this demonstrates is that in science, data are *generated* through experiment, observation, and measurement, then observed
- They are what is produced by *method*
- In other words: better described as
 - **capta** (“taken things”)
 - Not **data** (“given things”)



But in the humanities?

- Humanities is mostly **data** (“given things”)
- Humanities research is largely about the representation, analysis, and contextualisation of individual real world events, objects, people, or ideas
- Data is described, represented, and the subject of analysis much more than it is the product of experimentation or observation



Small, Thick, and Slow

- Often very small: **SMALL**
 - single paintings, single poems or words
- Extremely well curated: **THICK**
 - 100,000 words on 9-line poem
- Non-Rivalrous: **SLOW**
 - Reanalysed, re-represented, recontextualised over years
- E.g. Jane Austen Studies
 - 5 data points
 - 200 years of study



The work of science is
generating and analysing
capta (captured things)
through observation and
experiment

The work of Humanities
Research is representing
and analysing **data** (found
things that already exist)

Why does this matter?

1. **Although much Humanities research is (appropriately) “small, thick, and slow,” it is also, in theory, useful for “big capta” work**

- Collectively, traditional humanists produce a lot of *very* high quality data
 - *Intensely* curated datasets and data points;
 - Broadly compatible with each other (i.e. each generation reedits and reconsiders the canon);
- If we could find a way to capture the value of this traditional data in a way that would allow them to be reused,
 - We’d have extremely useful material to repurpose
 - We’d maximise benefit of traditional Humanities research

Why does this matter?

2. But FAIR infrastructure poorly adapted for “small data” projects

- The goal of traditional Humanities research is to publish data in context
 - Serve as primary sources for others
 - e.g. an edition of Jane Austen’s *Pride and Prejudice* is intended to support secondary work on that novel
 - Support very specific arguments about the instance in question
 - e.g. that there are three versions of *Hamlet*
- FAIR data requires a focus on data as stand-alone entity
 - Distinct PIDs (separate from analysis)
 - Data-centric, stand-alone metadata
- FAIR publication requires additional steps outside of standard workflow

Why does this matter?

3. We have no tradition (or infrastructure) for integrating FAIR data publication into traditional Humanities publication/research workflows

- Humanities research publishes data with analysis
 - Editions contain MSS photography, transcriptions, etc.
 - If you want the data, you have to access the edition
 - If the edition dies, you lose the (rich) data
- FAIR/Big data infrastructure has no tradition of supporting Humanities workflows
 - Repositories such as Zenodo, FigShare, even Humanities Commons treat data as solely stand-alone deposits
 - No system that supports integration in analytic contexts (e.g. as a source for or part of an edition)

So what to do?

A modern data project

- What we need is a workflow that encourages small-data researchers to prepare their datasets in a way that
 - respects their traditional requirements for the intensive curation and analysis of individual data points or small datasets
 - opens these **small, thick, and slow** datasets up to big data analysis
 - does not increase (and preferably reduces) the cost of production, publication, and maintenance
- **A workflow in which making “small data” ready for “big capta” is inherent in the workflow rather than a separate step.**

In other words

- Accept the traditional nature and use-case involved in the production and consumption of Humanities research data
 - I.e. recognise that FAIR must accommodate the small, thick, and slow as easily as it does the big stand-alone examples from STEM
 - Work with the traditional Humanities research workflow
- **As long as FAIR data publication means, in essence, publishing small, thick, and slow data twice (once in context and once without),**
 - **We will never fully reap the benefit of these important and potentially huge cultural datasets**
 - **Never improve the currently inefficient approach we take towards the publication of Humanities research data**

Part 2

Being FAIR to the small, thick, and slow

Introduction

- In this section we discuss the “Data-First” approach we are developing for the Visionary Cross Project
 1. The project and some of our parameters
 2. Background issues and models
 3. The implementation
 4. Further work

About the Visionary Cross Project

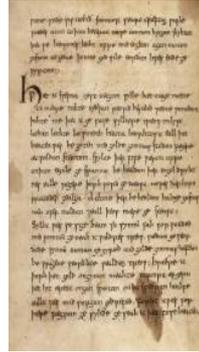
- 9 year-old SSHRC funded project to produce an “edition” and “archive” of the “Visionary Cross cultural matrix” in Anglo-Saxon England
 - “Edition” means “Scholarly mediated reproduction”
 - “Archive” means “dataset of facsimiles and transcriptions”
 - “Visionary Cross Cultural Matrix” means “Collection of individual objects that also belong together for cultural reasons”

About the Visionary Cross Project

- Objects include some of the best known objects and texts from Pre-conquest England and Scotland.

About the Visionary Cross Project

- Objects include some of the best known objects and texts from Pre-conquest England and Scotland.



Vercelli Book
Dream of the Rood
and Elene poems
(s. x/xi, South)



Ruthwell Cross
(s. viii, North)



Bewcastle Cross (s.
viii, North)



Brussels Cross
(s. x/xi, South)

About the Visionary Cross Project

- Interesting as individual objects and as a group:
 - Span period temporally, geographically, linguistically
 - (possibly) Earliest attested poetry
 - Complete runic poem
 - Include 1 of only 2~3 examples of poetic quotation
 - “Multiply attested” poetic text (>3% of the corpus)
 - Related to each other thematically (cult of the cross) and textually and/or artistically

About the Visionary Cross Project

- In other words we anticipate use as both
 - A traditional small-data project (as well as a not-so-traditional small-data project):
 - Individuals coming to us for limited amounts of data in the context of our thick description because they want to use our material as the primary source for subsequent work
 - A contribution to potential big-data purposes:
 - Data that can be used, reused, supplemented, and aggregated by others without negotiation

Project Requirements (pre-FAIR)

- A. Flexible:
 - Choose to view individual/group in appropriate format
- B. Extensible:
 - Add, rearrange, or reuse material without negotiation
- C. Authoritative:
 - Preserve credit/responsibility for all contributions
- D. Durable:
 - Permanently discoverable and available
 - Low/no maintenance

Different approaches over the years

- Wiki?
 - Flexible (e.g. categories/entries) (**A**)
 - Add and (re)connect material without negotiation (**B**)
 - But
 - Doesn't preserve Authority (**G**)
 - Requires ongoing maintenance (**D**)
 - One kind of presentation (**A**)

Different approaches over the years

- Game engine
 - Provided different ways of organising material and good at object/collection (**A**)
 - Preserved authority (**C**)
 - Some engines allowed some external contributions (**B**)
 - But
 - Requires others to use our system (**B**)
 - None strong on external contributions (**B**)
 - Requires ongoing maintenance (**D**)

OPenn (<http://openn.library.upenn.edu/>)

- Repository for MS information, images, transcriptions
- Replaced previous “turning the pages” interface
 - Open to machine access (i.e. via rsync, ssh, ftp, etc)
 - Human readability
- = a lightly-skinned directory structure (i.e. a RESTful-like API)
 - Human-readable HTML pages over an XML-based background and simple directories



The screenshot displays the OPenn website interface. At the top, it says "primary digital resources OPENN available to everyone" with navigation links for "Introduction", "Technical Help", "Repositories", "Curated Collections", and "Search". Below this, the "Home" page features "OPenn: Hosted Repositories" with a list of repositories including the Abraham Lincoln Foundation, Bryn Mawr College Special Collections, Christ Church, and the College of Physicians of Philadelphia. A table lists files with columns for "Name", "Last modified", and "Size".

Name	Last modified	Size
CuratedCollections.html	2019-07-06 18:20	4.5K
Datav	2019-03-20 12:04	-
ReadMe.html	2019-07-06 18:20	13K
Repositories.html		
Search.html		
TechnicalReadMe.html		
robots.txt		

To the right of the table, a message states: "This XML file does not appear to have any style information associated with it. The document tree is shown below." Below this message is a snippet of XML code:

```
<?xml xmlns="http://www.tei-c.org/ns/1.0">
<teiHeader>
  <fileDesc>
    <titleStmnt>
      <title>
        Description of University of Pennsylvania LJS
        16: Speculum historiale, Books 25-28
      </title>
    </titleStmnt>
    <publicationStmnt>
      <publisher>The University of Pennsylvania
        Libraries</publisher>
    </availability>
  </licence
target="http://creativecommons.org/licenses/by/
This description is ©2015 University of
Pennsylvania Libraries. It is licensed under
a Creative Commons Attribution License
version 4.0 (CC-BY-4.0
https://creativecommons.org/licenses/by/4.0/
For a description of the terms of use see
the Creative Commons Deed
https://creativecommons.org/licenses/by/4.0/
```

OPenn (<http://openn.library.upenn.edu/>)

- Love approach because it touches on all parts of vision
 - Flexible (i.e. **A**): can skin different groupings, focus on individuals or collections
 - Extensible (i.e. **B**): can extract from system
 - Authoritative (i.e. **C**): preserves authority
 - Durable (i.e. **D**): requires no software maintenance

OPenn (<http://openn.library.upenn.edu/>)

- But not perfect
 - Inflexible (i.e. **A**): Hierarchical data structure (can't have machine readable virtual collections)
 - Not extensible (i.e. **B**):
 - Additions/reorganisations require server access
 - Collections are “official” (entire libraries/fonds)
 - Not durable (i.e. **D**):
 - Publisher responsible for maintaining server
 - No persistent identifiers
- Not **FAIR** (as we now know)

Requirements (further points)

- E. Externally registered persistent identifiers
- F. Users need to be able to present alternatives/additions to our material inside or outside the same system
- G. Has to be “Publish-and-Forget”: once we are finished with it, it needs to be maintained by others

I.e. Has to be Findable, Accessible, Interoperable, and Reusable: FAIR

Our solution

- Use Zenodo and GitHub to create an OPenn-like data repository, while answering its lacunae
- A “Data-first” approach to publication that
 1. Is human and machine readable
 2. Preserves attribution
 3. Open to non-negotiated addition, reorganisation, reuse
 4. Uses standard, third-party-maintained, persistent IDs
 5. Maintained for free by others (requires no post-publication maintenance by the project)
- FAIR

Zenodo

- EU-funded OpenAire Data Repository
 - Hosted at CERN
 - Guaranteed by EU
 - Accepts “all research outputs from all fields of science”
 - Assigns DOIs to all submissions (“conceptual” and “record”)
 - Based on Invenio Digital Repository Engine
 - Excellent metadata and LOD capabilities

GitHub

- Code repository, version control, distribution system
- Used by millions for developing code-based projects
- Recently added ability to publish web-pages using Jekyll-based “GitHub pages”
- Based on Open Source Git
- But
 - Recently bought by Microsoft (it’s always been private)
 - Not archival (conditions of use allow for suspension of service for any reason at any time)

Interaction of Zenodo and GitHub

- **GitHub repositories can be archived in Zenodo**
 - Snapshots are deposited in Zenodo as Zipped directories
 - Given a Zenodo DOI and treated like any other record
 - Replaces GitHub's non-guarantee with Zenodo's permanent guarantee
 - Presentation (versions) are also citable research objects (FAIR data AND FAIR code)
 - Limitations are minimal:
 - If Github dies, the archived zip needs to be rehydrated on a different server
 - But if the data is atomic, this is a plug-and-play process

An example:

Cædmon's Hymn

- Originally CD-ROM (2005)
- Now online (2018)
- Code published using GitHub pages
 - <https://caedmon.seenet.org/>
 - <https://seenet-medieval.github.io/caedmonshymn>
- Code base preserved as Zenodo object (in all versions)

Cædmon's Hymn
A multimedia study, edition and archive

● Daniel Paul O'Donnell

With the assistance of Dawn Collins, Matt Van Egmond, Jon Lane, Catherine Larson, Angela Mlynarski, Asia Nelson, and Lyndon Simmons (2005) and Titilola Babalola Aiyegbusi and Gurpreet Singh (2018).

Version 1.1
Internet Reprint

TEI P4 SGML, and XHTML, 1.0 Transitional Conformant Edition

SEENET, A.8

Cambridge: D.S. Brewer, 2005
SEENET, 2018

DOI (code): [10.5281/zenodo.1198856](https://doi.org/10.5281/zenodo.1198856)
[How to cite this edition](#)

An example: *Cædmon's Hymn*

- Originally CD-ROM (2005)
- Now online (2018)
- Code published using GitHub pages
 - <https://caedmon.seenet.org/>
 - <https://seenet-medieval.github.io/caedmonshymn>
- Code base preserved as Zenodo object (in all versions)

The screenshot shows the Zenodo interface for a digital object. At the top, the Zenodo logo is on the left, a search bar in the center, and 'Upload' and 'Communities' on the right. The user 'singhg@uleth.ca' is logged in. The main content area displays the title 'Cædmon's Hymn: A multimedia study, edition, and archive internet edition. Version 1.1' with a 'Book' and 'Open Access' badge. Below the title, it lists the author 'Daniel Paul O'Donnell' and other contributors. A description follows, mentioning the online republication and the Cambridge edition. A 'Preview' section shows a file tree for 'caedmonshymn-v1.1.zip', listing folders like 'seenet-medieval-caedmonshymn-dee0786' and 'fonts' with their respective file sizes. On the right side, there are statistics: 86 views and 7 downloads, a 'Tweeted by 1' badge, and an 'OpenAIRE' logo. At the bottom right, there is a 'Publication date' section (April 21, 2018), a DOI (10.5281/zenodo.1226549), and 'Related identifiers' including the project website and GitHub repository.

Visionary Cross as Data

- Combining two systems allows us to publish a data-centric edition that is
 - Flexible
 - Extensible
 - Authoritative
 - Durable
 - Externally registered persistent IDs
 - Maintained by others

Heart is the Zenodo record

- Basic unit of edition (1 record = 1 datum)
- Provides machine readability, extensibility, persistence, and archiving
- *Also acts as document server for rest of edition

Zenodo record

- Human and machine readable metadata record + file(s)
- *Typed “additional identifiers”
- *Two kinds of DOIs:
 - “Conceptual” (latest)
 - “Version” (current)
- *RESTful files URLs
 - No link rot

The screenshot shows a Zenodo record page for 'Ruthwell Cross - 3D Model Ambient (15M)'. The page is dated April 24, 2018, and is associated with the 'The Visionary Cross' community. The record includes a DOI of 10.5281/zenodo.1490879 and a list of files: an XML record, a 3D model, and a 2D thumbnail. A preview window shows a 3D model of a cross against a blue background. The right sidebar displays 77 views and 15 downloads, and lists the publication date, DOI, keyword, and related identifiers.

zenodo Search Upload Communities singhg@uleth.ca

April 24, 2018 Dataset Open Access Edit

Ruthwell Cross - 3D Model Ambient (15M)

Singh, Gurpreet; O'Donnell, Daniel; The Visionary Cross Project

This is a High-Resolution Ambient 3D model the Ruthwell Cross, developed as part of the ongoing Visionary Cross project. This model has a 15M poly count.

This record contains:

- An xml record: `_Ruthwell_3DModel_Ambient_15M_Metadata.xml`
- A 3D Model of the Ruthwell cross: `cross_AmbientOcclusion_15M.ply`
- A 2D thumbnail: `Ruthwell_Cross00.png`

The DOI for this version of the record is 10.5281/zenodo.1490879. The DOI 10.5281/zenodo.1490878 always points to the latest version of this record.

This file in .ply format is best viewed using 3DHOP but can also be viewed using 3D rendering software like MeshLab.

For individual panels and other related material go to The Visionary Cross community at: https://zenodo.org/communities/the_visionary_cross/

Communities: The Visionary Cross Remove

77 views 15 downloads See more details...

Indexed in OpenAIRE

Publication date: April 24, 2018

DOI: 10.5281/zenodo.1490879

Keyword(s): Ruthwell Cross, The Visionary Cross Project, 3D Cultural Heritage, Digital Cultural Heritage, Anglo-Saxon, 3D Model, 3DHOP

Related identifiers: Documented by: 10.5281/zenodo.1490863

Communities: The Visionary Cross

License (for files): Creative Commons Attribution 4.0 International

Zenodo record

- Human and machine readable metadata record + file(s)
- *Typed “additional identifiers”
- *Two kinds of DOIs:
 - “Conceptual” (latest)
 - “Version” (current)
- *RESTful files URLs
 - No link rot

The screenshot shows the Zenodo record creation interface. At the top, there are three radio button options for access: "Embargoed Access", "Restricted Access", and "Closed Access". Below these is a "License" field with a dropdown menu set to "Creative Commons Attribution 4.0 International". A note below the license field states: "Required. Selected license applies to all of your files displayed on the top of the form. If you want to upload some of your files under different licenses, please do so in separate uploads. If you cannot find the license you're looking for, include a relevant LICENSE file in your record and choose one of the Other licenses available (Other (Open), Other (Attribution), etc.). The supported licenses in the list are harvested from opendefinition.org and spdx.org. If you think that a license is missing from the list, please contact us."

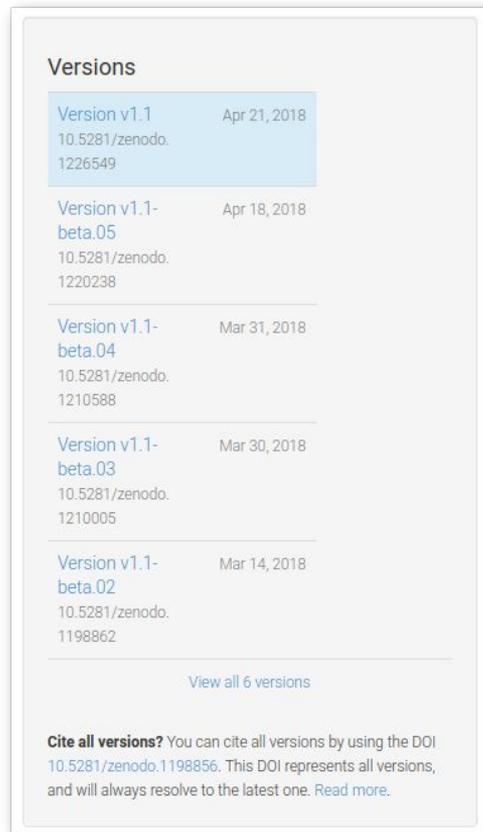
The "Funding" section is titled "recommended" and contains a text input field for the funding agency name, with a note: "Zenodo is integrated into reporting lines for research funded by the European Commission via [OpenAIRE](#). Specify grants which have funded your research, and we will let your funding agency know!". Below this is a "Grants" field with a dropdown menu set to "European Commission (EU)" and a text input field for the grant number. A note below the grants field states: "Optional. OpenAIRE-supported projects only. For other funding acknowledgements, please use the *Additional Notes* field. Note: a human Zenodo curator will need to validate your upload - you may experience a delay before it is available in OpenAIRE." There is a "+ Add another grant" button.

The "Related/alternate identifiers" section is titled "recommended" and contains a text input field for the identifier, with a dropdown menu set to "documents this upload". A note below the field states: "Specify identifiers of related publications and datasets. Supported identifiers include: DOI, Handle, ARK, PURL, ISSN, ISBN, PubMed ID, PubMed Central ID, ADS Bibliographic Code, arXiv, Life Science Identifiers (LSID), EAN-13, ISTD, URNs and URLs." There is a "+ Add another related identifier" button.

At the bottom, there is a list of optional fields: "Contributors", "References", "Journal", "Conference", "Book/Report/Chapter", "Thesis", and "Subjects", each with a right-pointing arrow.

Zenodo record

- Human and machine readable metadata record + file(s)
- *Typed “additional identifiers”
- *Two kinds of DOIs:
 - “Conceptual” (latest)
 - “Version” (current)
- *RESTful files URLs
 - No link rot



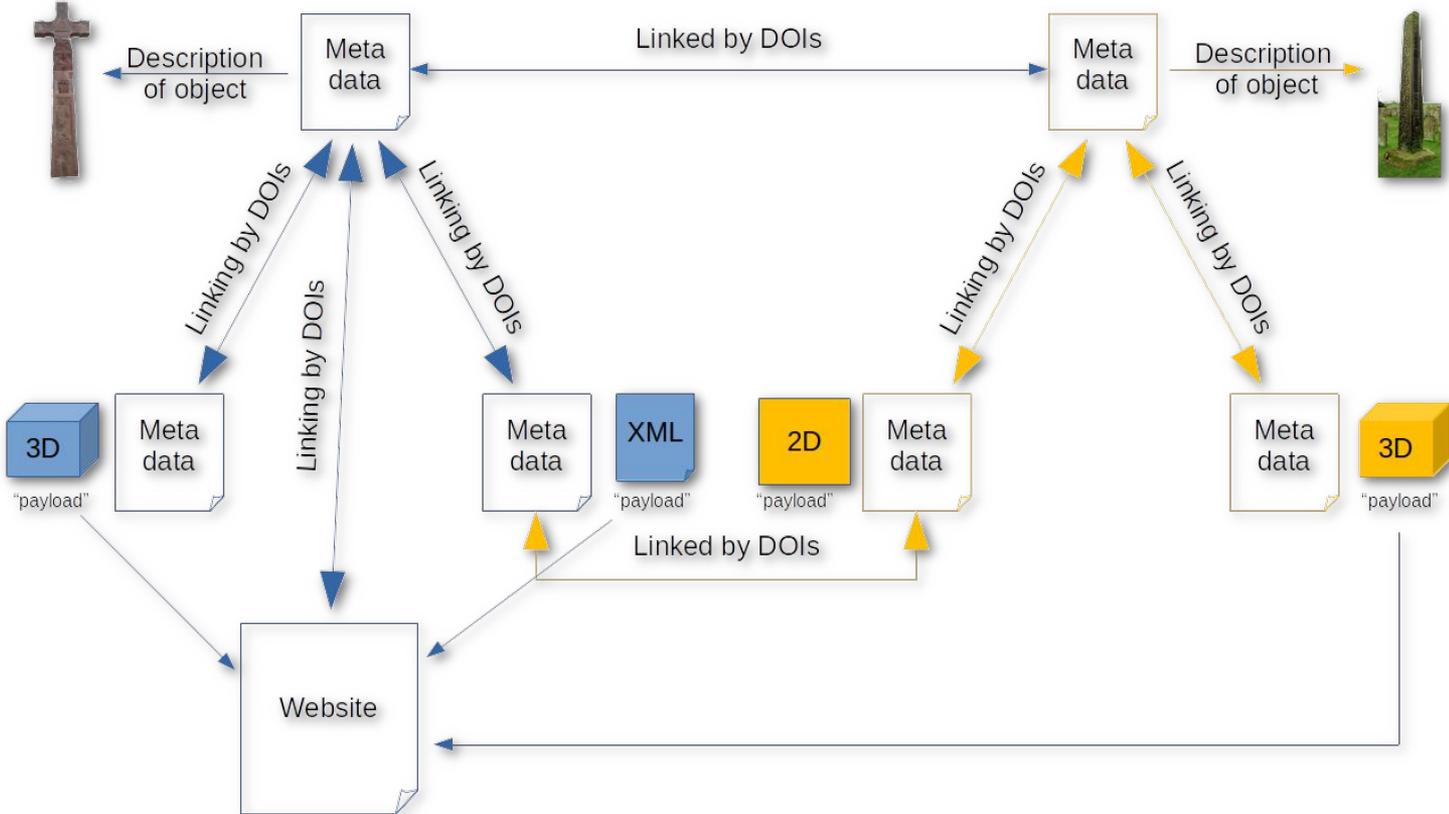
The screenshot displays a Zenodo record page titled "Versions". It lists five versions of a dataset, each with a version number, a date, and a DOI. The first version, "Version v1.1", is highlighted in blue. Below the list is a link to "View all 6 versions". At the bottom, there is a section titled "Cite all versions?" which explains that a specific DOI (10.5281/zenodo.1198856) can be used to cite all versions, and it will always resolve to the latest one. A "Read more" link is provided for further information.

Version	Date
Version v1.1 10.5281/zenodo.1226549	Apr 21, 2018
Version v1.1-beta.05 10.5281/zenodo.1220238	Apr 18, 2018
Version v1.1-beta.04 10.5281/zenodo.1210588	Mar 31, 2018
Version v1.1-beta.03 10.5281/zenodo.1210005	Mar 30, 2018
Version v1.1-beta.02 10.5281/zenodo.1198862	Mar 14, 2018

[View all 6 versions](#)

Cite all versions? You can cite all versions by using the DOI 10.5281/zenodo.1198856. This DOI represents all versions, and will always resolve to the latest one. [Read more.](#)

Edition is built around records



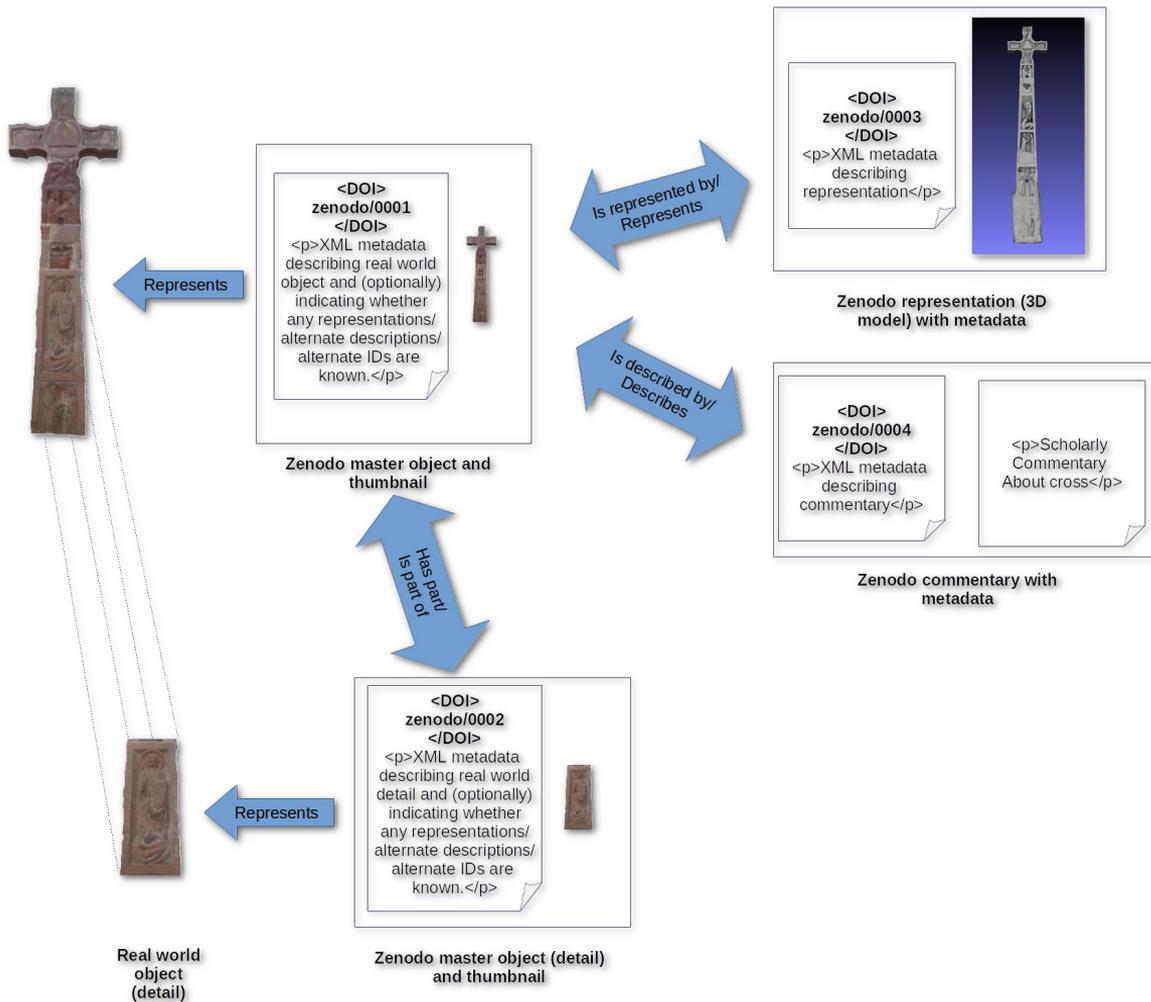


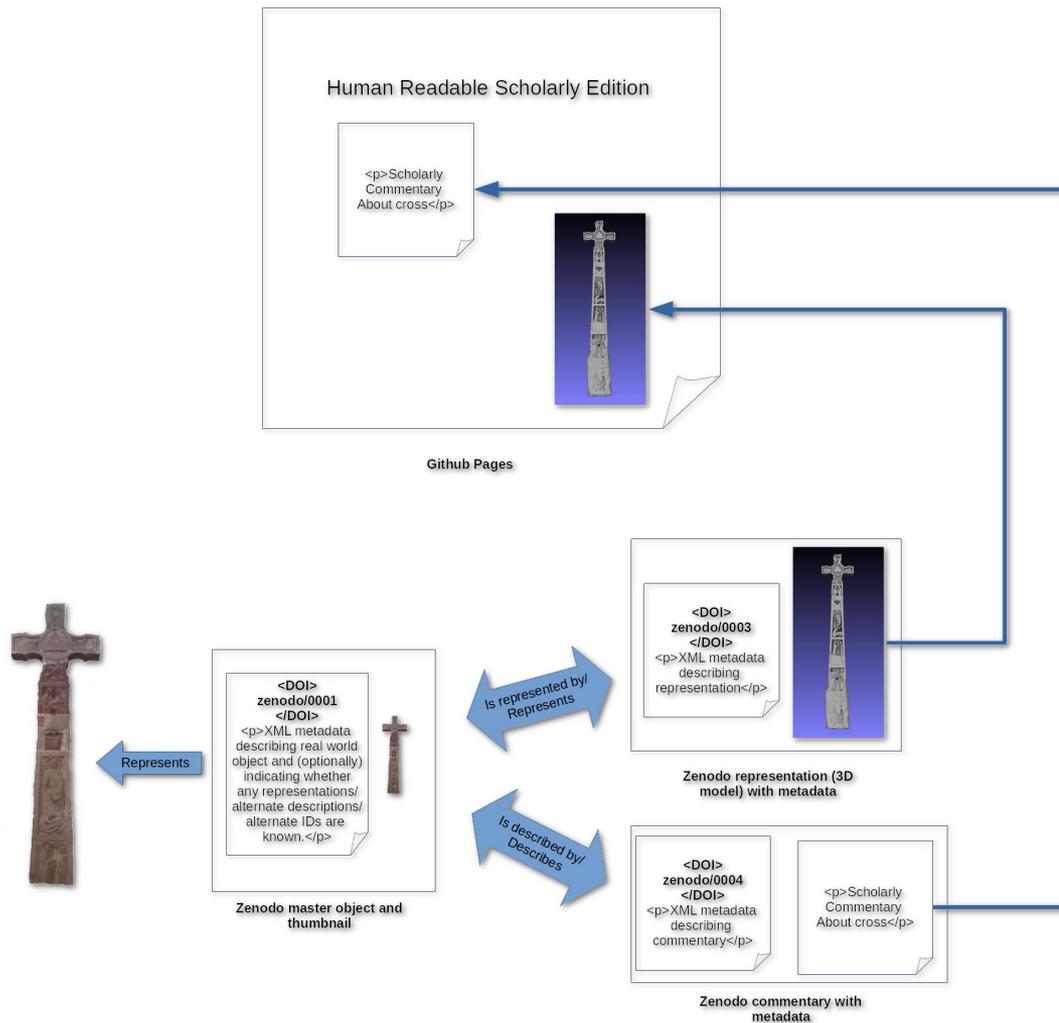
<DOI>
zenodo/0001
</DOI>

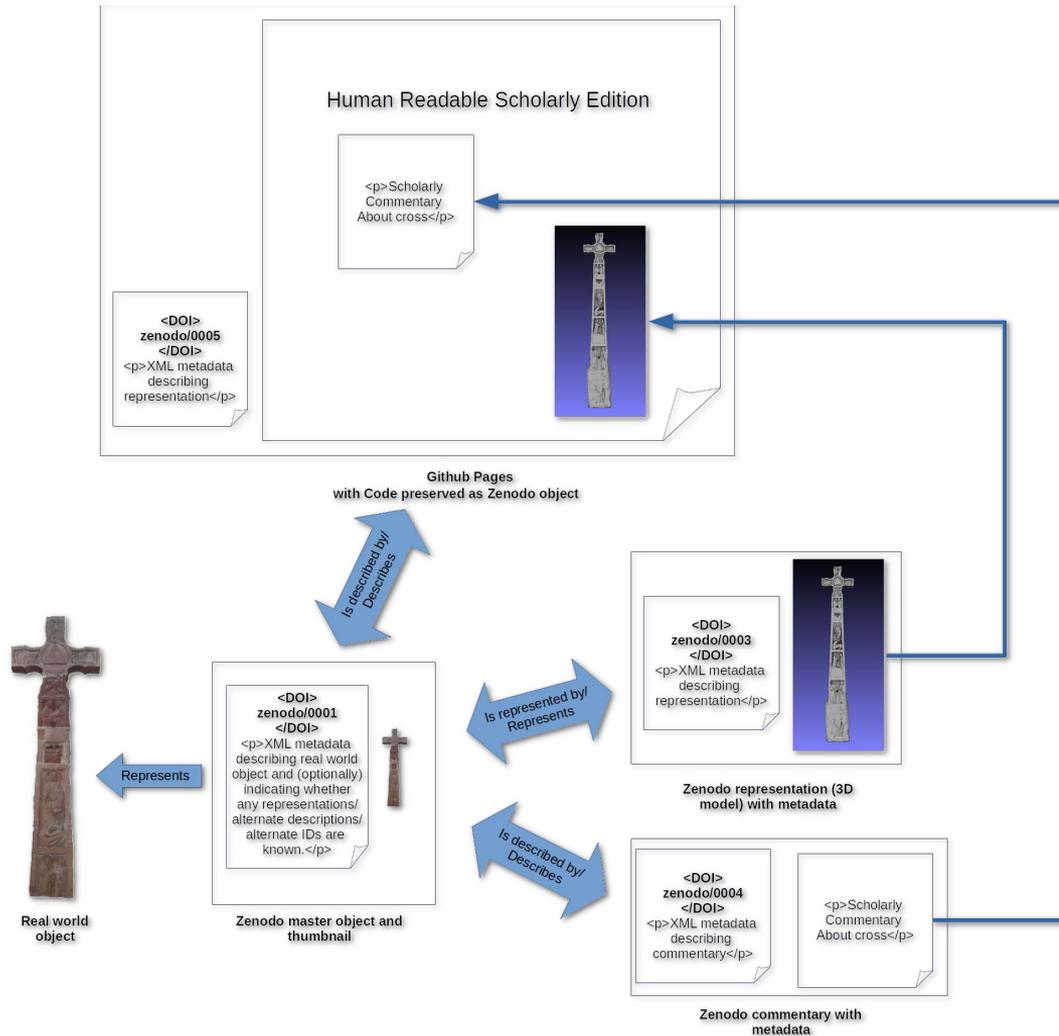
<p>XML metadata describing real world object and (optionally) indicating whether any representations/alternate descriptions/alternate IDs are known.</p>

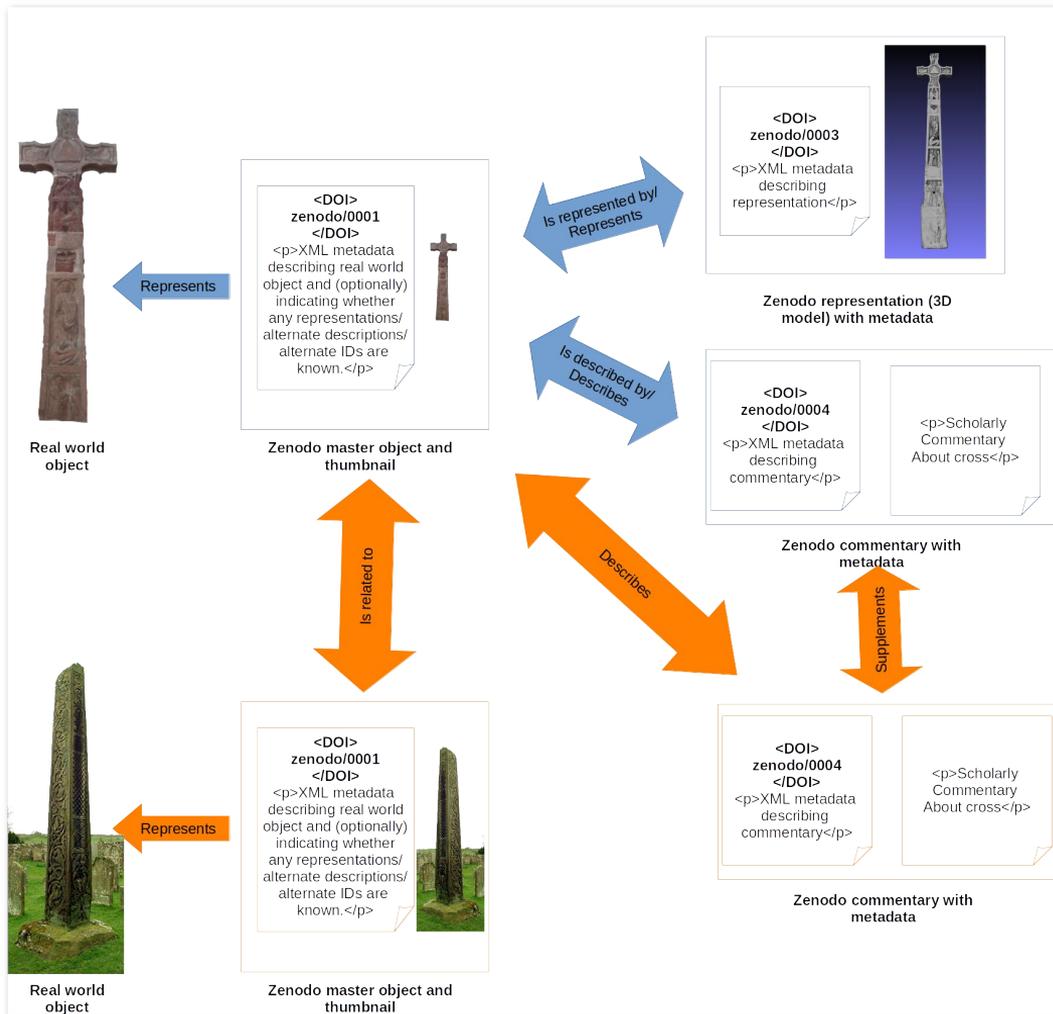


Zenodo master object and thumbnail









Advantages to this system

- Like OPenn
 - Human and Machine Readable
- Improve on OPenn
 - Persistent IDs (can be used RESTful)
 - FAIR
 - Not restricted to hierarchical arrangement or read only
 - Can be exported to variety of standards
 - Can be added to or rearranged by others
 - Maintained by archival specialists (i.e. commitment to preservation)
- **Supports small, thick, and slow publication in a FAIR format**

Disadvantages

- What is interesting about this approach is that it is accidental
 - While most features are supported,
 - Not all are (e.g. arbitrary ontologies)
 - Those that are are inconsistent across repositories (e.g. streaming; typed identifiers)
 - Support is often tentative or inadvertent
 - Conceptual vs Record DOIs
 - Restful DOI-based API
- While the ability to support Humanities data is there, the systems have not been designed with Humanities data in mind
- Supporting small, thick, and slow data is something that can be accommodated with relatively little work

Part 3

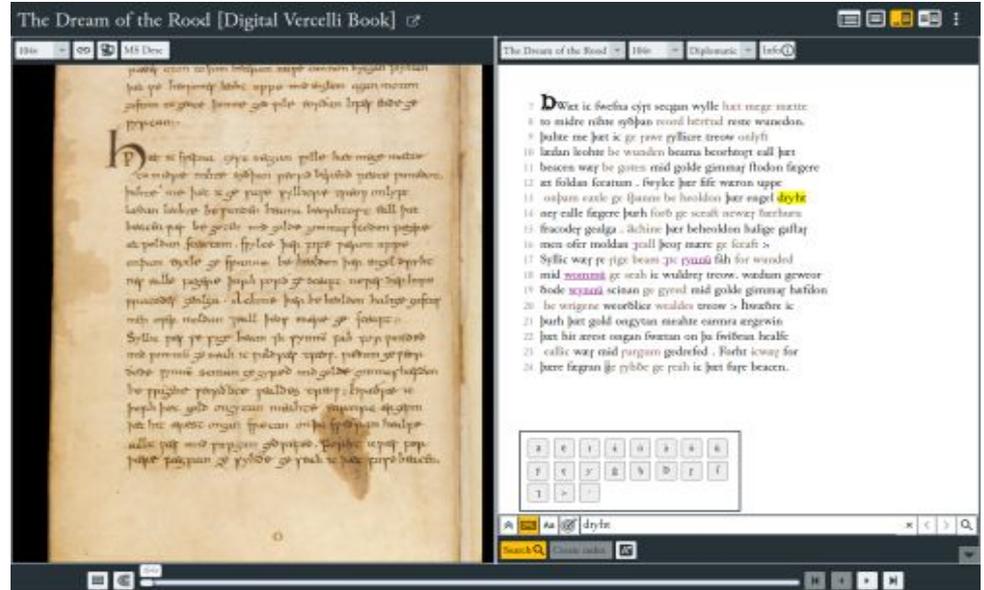
Providing assistance to the small, thick,
and slow: The case of the EVT

EVT - Edition Visualization Technology

- An open source tool to publish TEI-based digital editions
- Born as the browser component of the Digital Vercelli Book project
- Initial needs:
 - a viewer for manuscript images together with TEI diplomatic transcription, image-text linking, two edition levels, text search
 - simple to use and deploy, Web standards-compliant
- Designed from the start to be a modular tool, so that it could evolve in a general purpose tool useful for other projects
- Collaboration with other projects led to many new features, now aiming at covering the most common needs for TEI XML-based editions:
 - digital facsimile
 - diplomatic / interpretative edition together with manuscript scans
 - critical edition

EVT and the Visionary Cross project

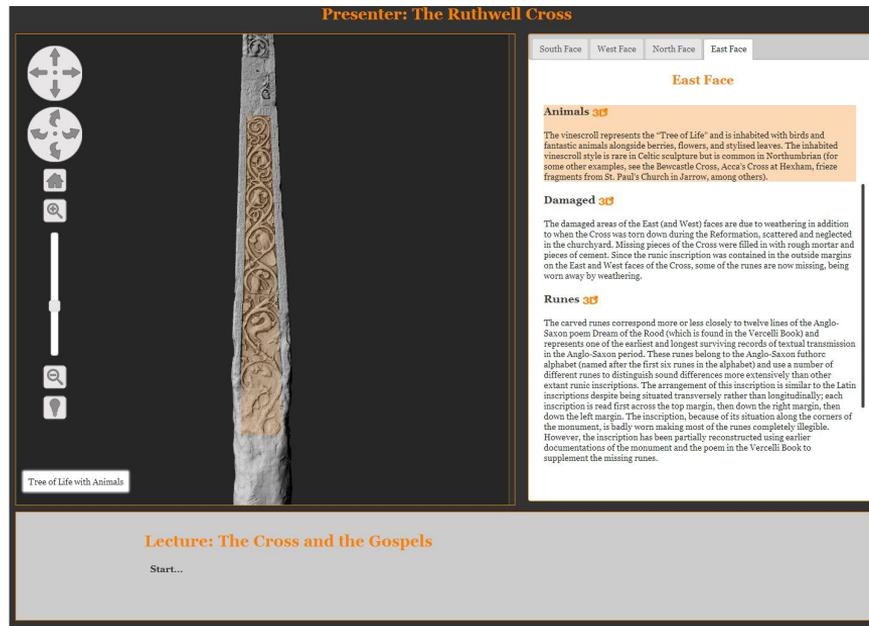
- The original plan:
 - EVT to publish the Vercelli Book text of *The Dream of the Rood*
 - a new tool to publish the 3D models and related transcriptions / commentary text



EVT 2: *The Dream of the Rood*

EVT and the Visionary Cross project

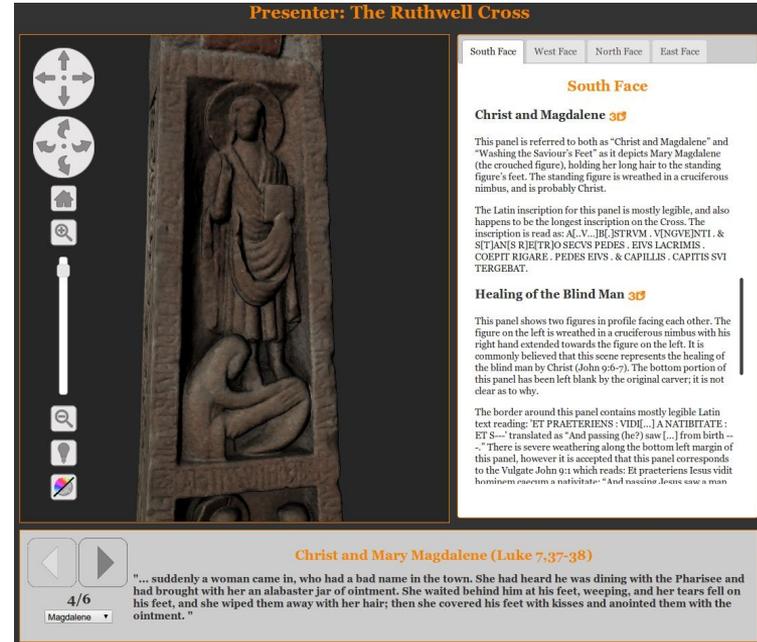
- The original plan:
 - EVT to publish the Vercelli Book text of *The Dream of the Rood*
 - a new tool to publish the 3D models and related transcriptions / commentary text
 - → The Ruthwell Cross Presenter developed by CNR researchers



The Ruthwell Cross Presenter

EVT and the Visionary Cross project

- The original plan:
 - EVT to publish the Vercelli Book text of *The Dream of the Rood*
 - a new tool to publish the 3D models and related transcriptions / commentary text
 - → The Ruthwell Cross Presenter developed by CNR researchers

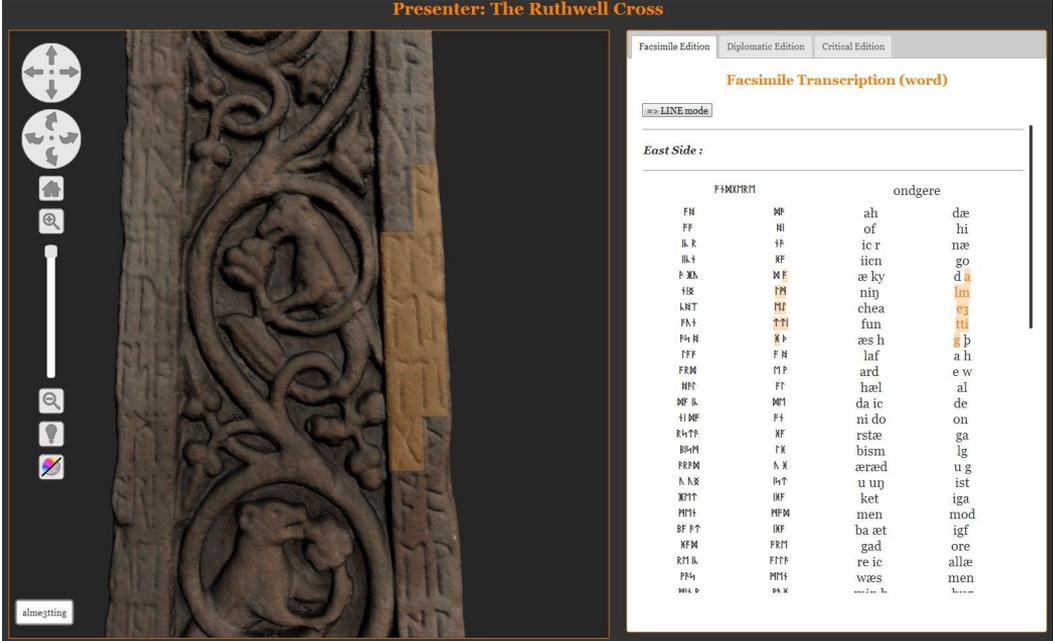


The Ruthwell Cross Presenter

EVT and the Visionary Cross project

- The original plan:
 - EVT to publish the Vercelli Book text of *The Dream of the Rood*
 - a new tool to publish the 3D models and related transcriptions / commentary text
 - → The Ruthwell Cross Presenter developed by CNR researchers

Presenter: The Ruthwell Cross



Facsimile Edition | Diplomatic Edition | Critical Edition

Facsimile Transcription (word)

=> LINE mode

East Side :

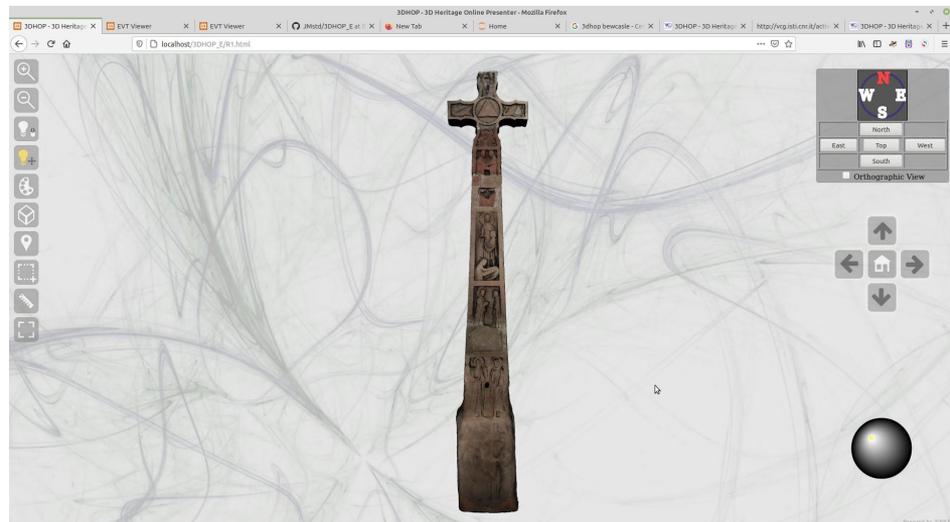
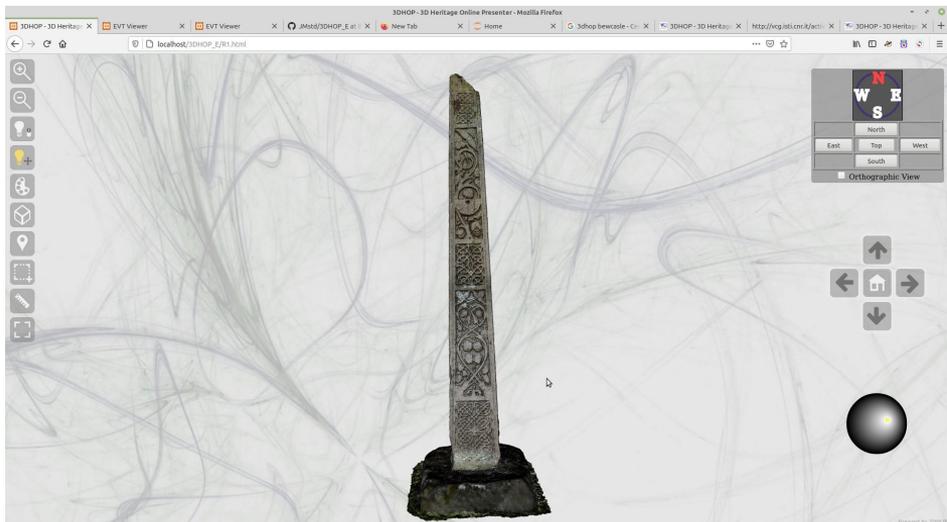
FAMOTUM		ondgere	
FN	MF	ah	dæ
FF	NI	of	hi
IK R	IF	ic r	næ
IK f	XF	iien	go
F 3X	MF	æ ky	d a
IF	FM	nij	lm
LNT	MF	chea	eg
FA f	FT	fun	tti
FM N	X P	æs h	g p
FFM	F N	laf	a h
FRM	MP	ard	e w
MF	FT	hæl	al
MF IK	MF	da ic	de
IF MF	F f	ni do	on
IK f F	XF	rstæ	ga
IK f M	FX	bism	lg
FRM	A X	ærad	u g
A A X	IF f	u unj	ist
MF f	IK F	ket	iga
MF f	MF M	men	mod
BF f	IK F	ba æt	igf
XF M	FK F	gad	ore
FM IK	FF f	re ic	allæ
FM f	MF f	wæs	men
MF f	FX X

almegting

The Ruthwell Cross Presenter

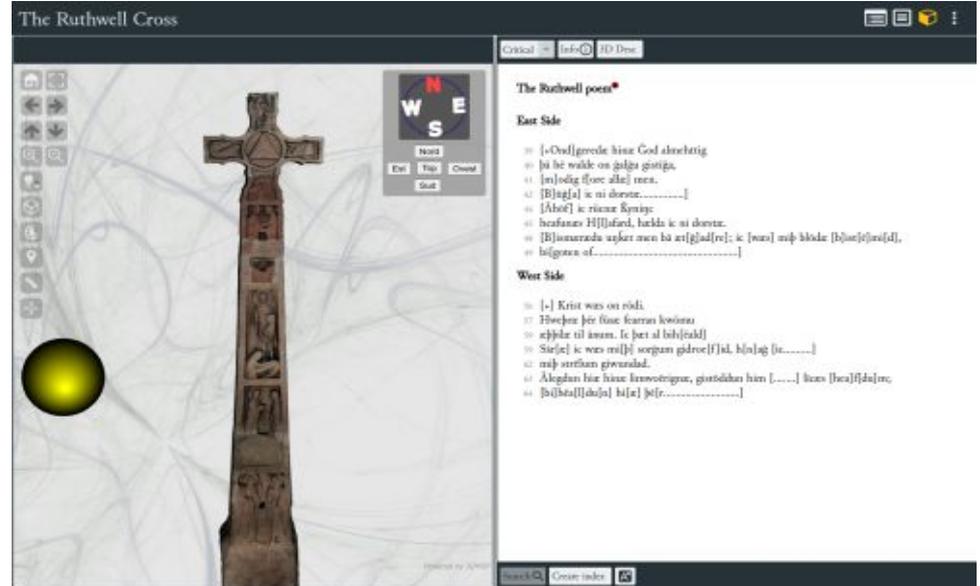
EVT and the Visionary Cross project

- The successor to the Ruthwell Cross Presenter:
 - **3DHOP** a powerful 3D viewer again developed by CNR researchers
 - stable and feature-rich, but very limited text capabilities



EVT and the Visionary Cross project

- The new plan:
 - integrate 3DHOP in EVT to publish both 3D objects and related text
 - make a better connection between 3D and text(s) and ms images than in the Ruthwell Cross Presenter



Experimental 3DHOP integration in EVT 2

A good plan, but ...

- It still didn't answer (some of) the crucial questions we discussed above:
 - How do we put everything together?
 - We always intended for our “multi-object edition” to be expandable, how to best make that possible?
 - How to handle data that could go from a few Kilobytes (single text metadata) to many Gigabytes (3D models)?
 - How to make our data available for other projects, and ensure that it stays available in the long term?
 - How do we make our data FAIR?

EVT and the distributed edition

- What I realized is that, in spite of my intentions, EVT was still very much grounded in the “printed page paradigm”, at least with regard to how the edition is designed and implemented
- A monolithic object which may be able to access (better: to link) to external resources, but still would keep all of the core components (text, images, 3D models) on the publication server
- The idea of “opening up” an EVT edition became part of the wider methodological thinking we were doing about how humanities data dissemination works (or doesn’t work)

EVT and the distributed edition

- Actually an EVT goal dating back quite some time
- Only recently all the necessary technologies seem to be falling into place
 - LOD repositories
 - IIF framework for images
 - CTS and DTS protocols for text
 - Zenodo as a safe, long term general purpose repository
 - GitHub as a server of live pages
- To use those effectively we need to work both on the general methodology and to experiment with the existing tools
- Great opportunities, but also some risks and complications (what if PIDs aren't really *persistent*?)

EVT and the distributed edition

Leges Langobardorum
project:

- accessing images available on IIF-compliant servers
- e.g. the Codex Sangallensis 730 on the e-codices digital library

The screenshot displays the digital edition of the Edictum Rothari. The interface is split into two main panes. The left pane shows a high-resolution image of a manuscript page, featuring a large, ornate initial 'E' in red and blue, followed by Latin text in a Gothic script. The right pane shows a digital edition of the text, with numbered annotations (e.g., 01, 02, 03) and a search bar at the top. The annotations appear to be cross-references or commentary on the text. The overall layout is clean and professional, typical of a digital library interface.

Page → IIF full image

<pb facs="https://www.e-codices.unifr.ch/loris/csg/csg-0730/csg-0730_020.jp2/full/full/0/default/jpg/"/>

Long term goals for EVT and Visionary Cross

- Continue developing 3DHOP and EVT integration → crucial feature: hotspots on the 3D model linked to the transcription text
- General functionality to load edition data from long term repositories such as Zenodo (also GitHub pages for live updating, already working with EVT 2)
- General compliance with major image / text serving protocols over the Internet
- Better integration of LOD into any EVT-based edition
- Exposing the edition data so that they can be “distant processed” (this could be done even without implementing an *ad hoc* API)
- Exposing the edition data so that they can become LOD (this requires several other processing resources, e.g. being able to extract RDF-like triples)

Part 4.

Next steps

“Good things come in small packages”

Next steps

- Formalise this use case and feature-set
 - Build a prototype publication system within Zenodo/Github
 - Identify (and develop) required features where they are tentative or missing
 - Test system out on existing publications and data
- Encourage others to look at their data publication in this atomic/FAIR way
- “Good Things Come in Small Packages”: a SSHRC Partnership Development Grant/Community of Practice for people and institutions interested in the FAIR publication of Small Data.

Contact us!

daniel.odonnell@uleth.ca
roberto.rosselidelturco@unito.it

Questions?

daniel.odonnell@uleth.ca
roberto.rosselidelturco@unito.it