Report on the FLOSS Task Force

12/1/2016

Task force participants

1. Vladimir Alexiev (Data and Ontology Management, Ontotext Corp)
2. Agiatis Benardou (Digital Curation Unit, Athena Research Centre)
3. Valentine Charles (Europeana Foundation) Task Force Co-chair
4. Nephelie Chatzidiakou (Digital Curation Unit, Athena Research Centre)
5. Panos Constantopoulos (Digital Curation Unit, Athena Research Centre)
6. Themistoklis Karavellas (Netherlands Institute for Sound and Vision)
7. Sergiu Gordea (AIT - Austrian Institute of Technology GmbH)
8. Ana Isabel González Sáez (MECD - Ministry of Education, Culture and Sports. Spain)
9. Gregory Markus (Netherlands Institute for Sound and Vision) Task Force Chair
10. Lyndon Nixon (MODUL, University Vienna)
11. Leonidas Papachristopoulos (Digital Curation Unit, Athena Research Centre)

1. Introduction

The EuropeanaTech FLOSS Inventory Task Force ran from June 2015-December 2015. It consisted of 17 members from across Europe involved in various areas of work within digital humanities and cultural heritage.

What is the FLOSS Inventory?

The FLOSS Inventory[1] is a list of Free, Libre, Open Source Software relevant for the digital cultural heritage sector at large. It was started during Europeana v 2.0 by the Netherlands Institute for Sound and Vision. It currently (following this Task Force) contains 218 items.

Why the FLOSS Inventory?

The goal during Europeana v 2.0[2] and the creation of the FLOSS Inventory[3] was for EuropeanaTech to take an active role in improving the standing of software available. This meant collecting, organizing, assessing, and working closely with developers to improve existing documentation and tools.

---

[1] http://bgweb.nl/floss/

[2] http://pro.europeana.eu/files/Europeana_Professional/Projects/Project_list/Europeana_Version2/Deliverables/Ev2%20D7.2%20Report%20of%20Inventory%20of%20FLOSS%20Documentation%20and%20Sustainability.pdf

[3] http://pro.europeana.eu/taskforce/europeanatech-floss

2. Goals of Task Force

The goals of the Task Force as stated in its charter were quite narrow in scope and focused primarily on being able to structure and enrich the Inventory. After several meetings it became clear that some of these desired goals would require more resources and time as well as a set of additional expertise among the Task Force participants. However, the Task Force capitalized on the strengths and knowledge of its participants.

Below in italics is the original Task Force proposal and with goals and expected work.

Enrichment and Improvement

As per the request of the European Commission, and from the inherent necessity to make the FLOSS Inventory something more than a simple list of tools, this Task Force investigates and proposes several ways to improve the current FLOSS Inventory. This will be done in several ways:

a. *Standardizing necessary metadata elements to describe the tools : not every (descriptive) properties are necessary, or are updated on regular basis or pertinent to the users. The Task Force will assess which metadata information is most important and easy to maintain as to not have a plethora of out-dated information on the repository.*

b. *All the fields for existing tools and new tools will be updated as thoroughly as possible, most importantly, checking for dead or broken links.*

c. *List cleanse: any tools that have been inactive for several years and/or contain no documentation and/or have no functional demo will be removed. This is to maintain a certain level of quality over quantity within the inventory.*

d. *Reviews:* Task Force *members were asked to review FLOSS entries. The reviews cover the following aspects: Quality of Documentation, Ease of Adaptation, Existence of functional installations/demonstrations, Access to source code .*

e. *A similar directory has been developed for Digital Humanities called the DiRT Directory[4]. The Task Force will assess the overlap between the two directories and investigate directions for collaboration.*

---

[4] http://dirtdirectory.org/

Gathering new software and tools

Currently the FLOSS Inventory has 307 entries[5]. While the list covers many of the paramount technical aspects of digital cultural heritage, we believe that more software is open and freely available, even from within the Europeana Network.
As part of this Task Force we would like to create a critical mass of aggregation. The difficulty of this simply lies with the amount of contributors to the inventory.

While quantity is an easy measure of success, the FLOSS Inventory will not have a KPI based on the number of tools, but rather success and quality will be measured based on how well represented each aspect of OSS relevant for digital humanities and digital cultural heritage is.

All new entries will be required to have reviews.

Improvement of software taxonomy

Cultural heritage software is diverse. Many aspects of software and informatics are relevant for institutions and developers in the sector. From tools that enable the re-use of metadata or content, backend Collection Management Systems (CMS), cataloguing software, to front-end exhibition and presentation systems. The field is extremely diverse and vast. Due to this broad scope, it is imperative that softwares added to the inventory fit into a proper taxonomy to improve findability of relevant tools. This section will also take into consideration the Digital Humanities taxonomy, TaDiRAH[6] used by the DiRT Directory.

The outcomes of the Task Force will primarily be visible in the Inventory itself. An overview of the process and selection of new tags and classifications of DCH OS tools will be drafted.

As you will see in this report, the Task Force has achieved some of these goals, left others to be saved for a later date, and accomplished others that were not originally planned.

---

[5] As per June 2015
[6] https://github.com/dhtaxonomy/TaDiRAH

3. User profiles

One of the first tasks the Task Force undertook was creating user profiles to make a proper analysis and better understand the needs of the stakeholders that use the FLOSS Inventory.

First we had to solve the initial challenge: understanding what the Inventory is. The Inventory is not a clear-cut list of tools and software for experienced developers nor is it filled with software for marketers, curators, and other non-technical people in cultural heritage. Therefore, because of its diversity, determining who can use it as a reference list meant first understanding exactly what the Inventory contained and how wide the scope was.

To accomplish this, each Task Force participant was asked to examine a selection of tools and hypothesize use cases for each entry. Doing so meant that we could use the FLOSS Inventory as our starting and end point. After completing this activity, the Task Force could start defining some user profiles.

Below you can see several examples of user profiles the Task Force has defined. The full list is available in Annex 1.

Proto-Persona	Art researcher
Age	40
Skill Level (Years developing, knowledge)	10 years in specific domain, 1 in IT
Sector (Museum, Library, creative, etc)	Visual arts (re future Europeana Art channel)
Desired outputs	Semantic search and classification to discover and classify objects related to art.

Proto-Persona	Curator in small (local) heritage institution
Age	40
Skill Level (Years developing, knowledge)	Art historian, engaged in promotion of history of his hometown, IT fan, recently set up his blog by installing wordpress on a virtual hosting facility (using precise step by step tutorial)
Sector (Museum, Library, creative, etc)	Museum/archive
Desired outputs	Something which will help him to promote his collections on-line - maybe a basic digital library system or some CMS popular in heritage sector.

Proto-Persona	Digital Librarian
Age	25+
Skill Level (Years developing, knowledge)	Degree in library science or humanities and information technology or familiar with information technology
Sector (Museum, Library, creative, etc)	Library or Museum or Software vendor
Desired outputs	Cleaned datasets, structured textual data (eg. TEI files), metadata mapping, optical character recognition, data or metadata conversion, bulk processing, schema definition.

Proto-Persona	Student
Age	21
Skill Level (Years developing, knowledge)	1.5 years

Sector (Museum, Library, creative, etc)	Creative
Desired outputs	Student trying to find platform that will allow for live video annotations and way for the annotations to be exported and categorized for a research paper on crowd sourcing.

As can be seen, there is still a great amount of diversity between potential users. However, this is the nature of digital cultural heritage. Technical skill sets vary as greatly as the aspects of the technical work carried out within the sector do. In addition, the overlap with digital humanities makes singling out one type of user within the Europeana Network Association extremely difficult. We conclude that FLOSS primarily targets individuals with a technical background and some knowledge of programming but is equally relevant for highly experienced developers as well as those who have the most basic knowledge of the Internet.

3.1 Dealing with scope and diversity

Maintaining a broad scope of openly licensed software and tools relevant for all areas of DCH available on the Inventory is a smart decision for several reasons:
1. Avoid duplication of work
2. Save time and money for small (or large) institutions
3. Encourage transnational collaboration to build upon existing tools instead of developing as institutional islands
4. Develop sense of community
5. Smarter usage of public funding

In other words, before undertaking a product development process or submitting a project proposal for software and tool development, institutions should check the FLOSS Inventory to see whether or not someone already has or currently is developing something similar.

Since the FLOSS Inventory is potentially of interest for so many different types of users with different knowledge levels, a Google Spreadsheet is neither an appropriate host for it, nor an appropriate tool for publishing documentation of tools. Making the Inventory understandable, available and appealing for all will require concise metadata fields, quality metadata, clear vocabularies, as well as storage and display through a well designed CMS. These aspects are all presented and discussed in the following sections.

4. Aligning and structuring FLOSS Inventory

For the EuropeanaTech FLOSS Inventory to be representable in a standardized way, the NeDiMAH Methods Ontology (NeMO) was used to provide Activity Type designations for each tool included in the Inventory. NeMO is an ontological model of scholarly practice in the arts and humanities, developed under the work plan of the ESF Research Network NeDiMAH. It integrates existing taxonomies of scholarly methods and tools, such as TaDIRAH, the arts-humanities.net and Oxford taxonomies of ICT methods, DHCommons, CCC-IULA-UPF and DiRT, through appropriate mappings of the concepts defined onto a semantic backbone of NeMO concepts. It thus enables combining documentary elements on scholarly practices of different perspectives and using different vocabularies.

In the context of the EuropeanaTech FLOSS Inventory Task Force, NeMO was used in order to align the tools included in the inventory with Activity Types integrated in the NeMO ontology. This process responds to the need of standardization of the content of the FLOSS Inventory and permits both its alignment with a rigorous taxonomy of research activities and its potential link to a model describing the overall scholarly work in the arts and humanities. Furthermore, NeMO is a CIDOC CRM - compliant ontology that can be represented in both document and machine readable forms.

The NeMO Activity Types are dispensed around five core activities corresponding to the main steps of the scholarly research lifecycle, namely acquiring, communicating, conceiving, processing and seeking (see Figure 1).
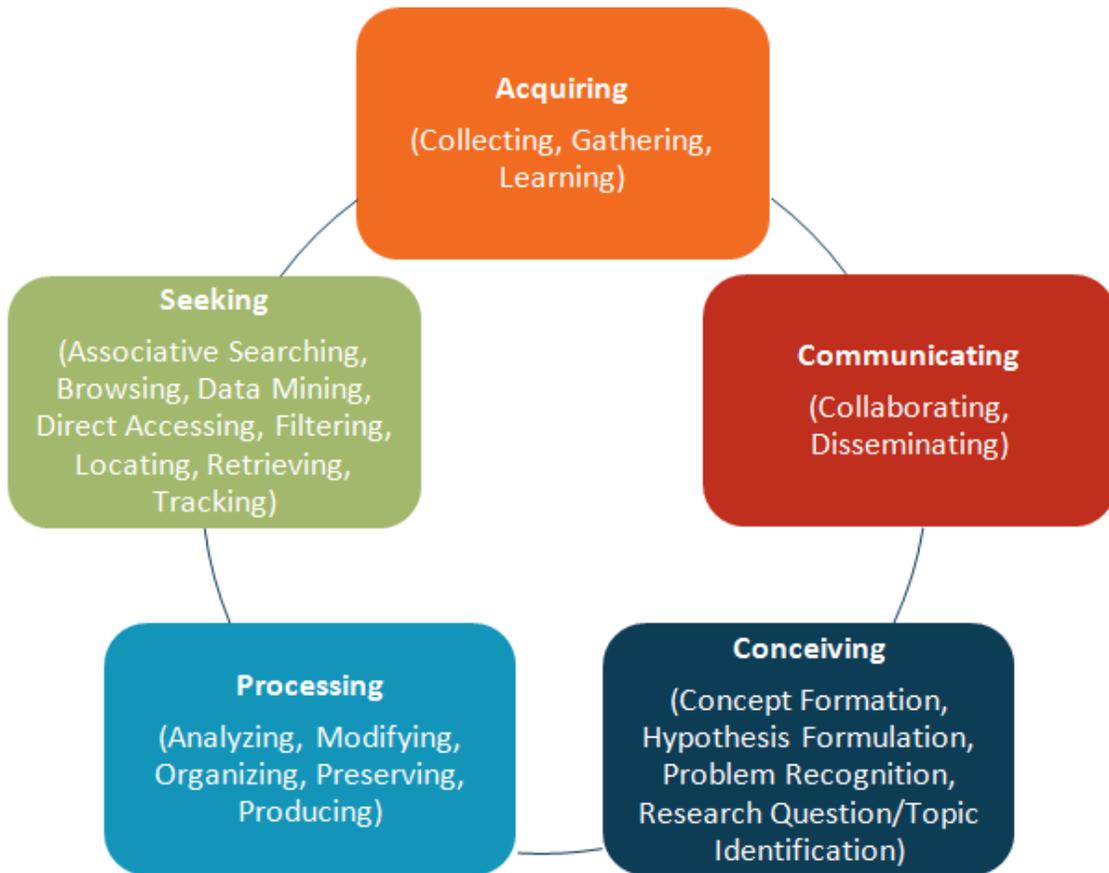
Figure 1. The core NeMO Activity Types

These activities comprise sub-activities and sub-sub-activities for as many levels as needed to fully grasp the network of a specific activity. Tools in the FLOSS inventory have been assigned one or more Activity Types, depending on their functions and outcomes. This categorization permits a grouping of the available tools according to the function they serve. Taken as a whole, one can observe that most tools[7] (72%) are related to the Processing Activity Type, while fewer tools are related to Seeking (13%) and Communicating (11%) and very few or none at all are related to Acquiring (4%) and Conceiving (0%) - see Figure 2.

---

[7] *As most tools were assigned more than one Activity Types, the percentages presented here do not refer to the number of tools but to the sum of Activity Types assigned to the tools.*
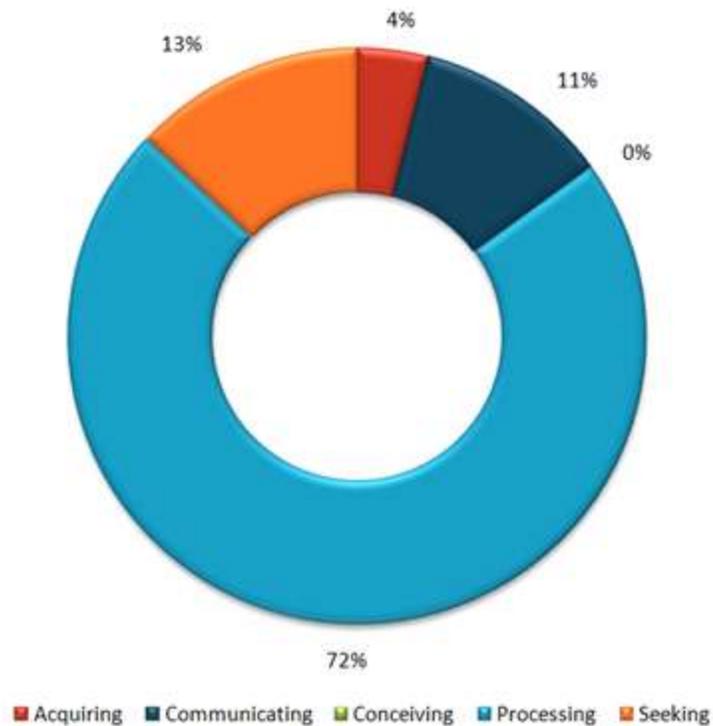
Figure 2. Distribution of the FLOSS tools according to the NeMO Activity Types

Focusing on a lower level, the tools related to Acquiring have functions such as data extraction, metadata extraction, gathering, harvesting and collecting (see Figure 3). Tools related to communicating are mostly relevant to publishing and resource sharing, while others are associated with activities such as collaborating, crowdsourcing, networking and consulting (see Figure 4). As mentioned above, most tools are related to the broad activity of processing. Most of these tools are associated to managing, visualizing, annotating, adding meta-information, preserving and archiving, while a significant number of tools are related to activities such as data recognition, modifying, web developing, editing, parsing and producing (see Figure 5). Finally, Figure 6 showcases the activities related to seeking. Most of these tools are used for browsing, seeking and extracting data, while others are associated to activities such as direct accessing, locating, retrieving, extracting metadata, data mining, filtering, aggregating and harvesting.
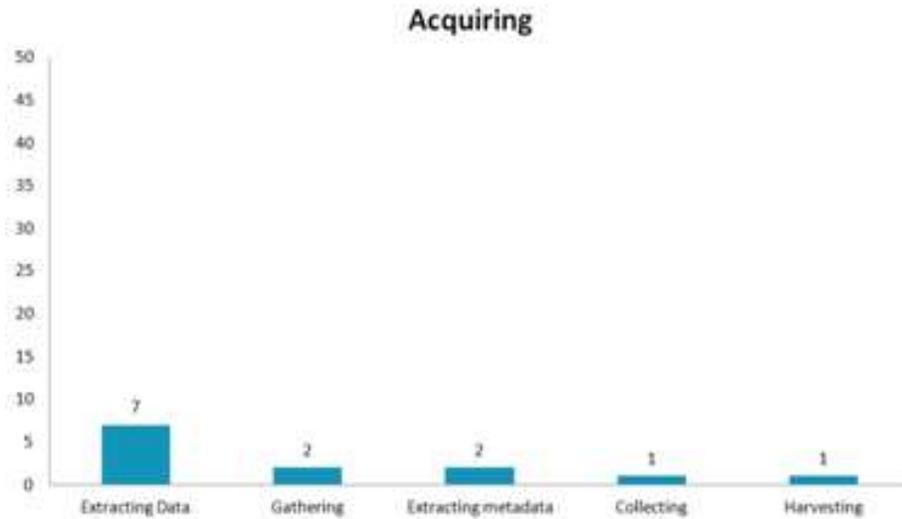
Acquiring

Figure 3. Distribution of the FLOSS tools according to the lower-level activities of the Activity Type "Acquiring".
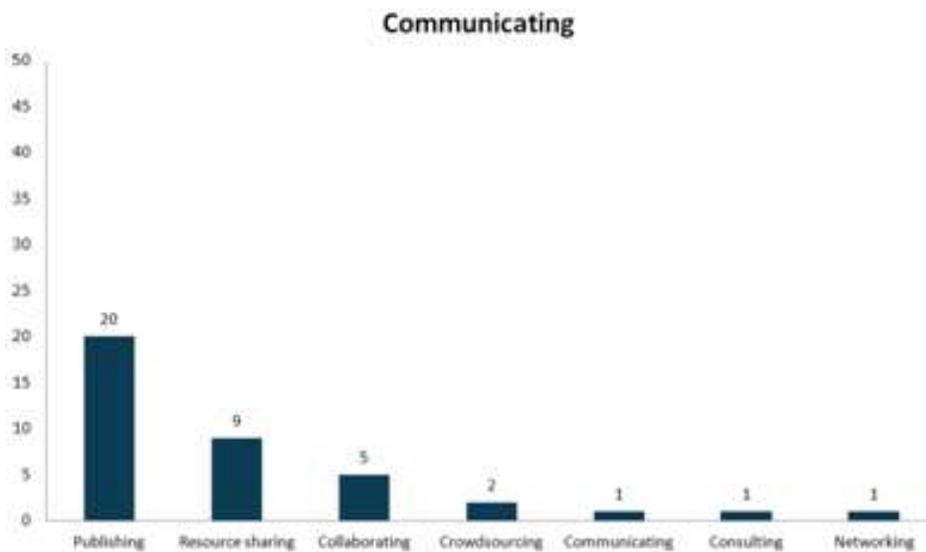
Communicating

Figure 4. Distribution of the FLOSS tools according to the lower-level activities of the Activity Type "Communicating".
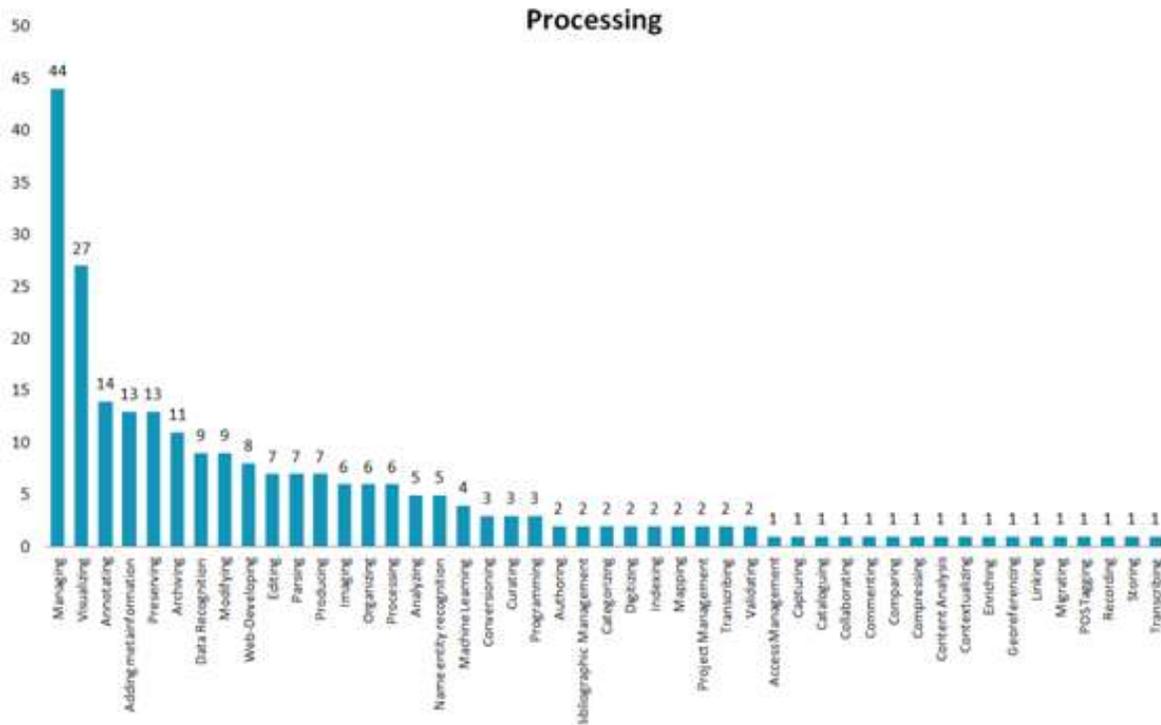
Figure 5. Distribution of the FLOSS tools according to the lower-level activities of the Activity Type "Processing".
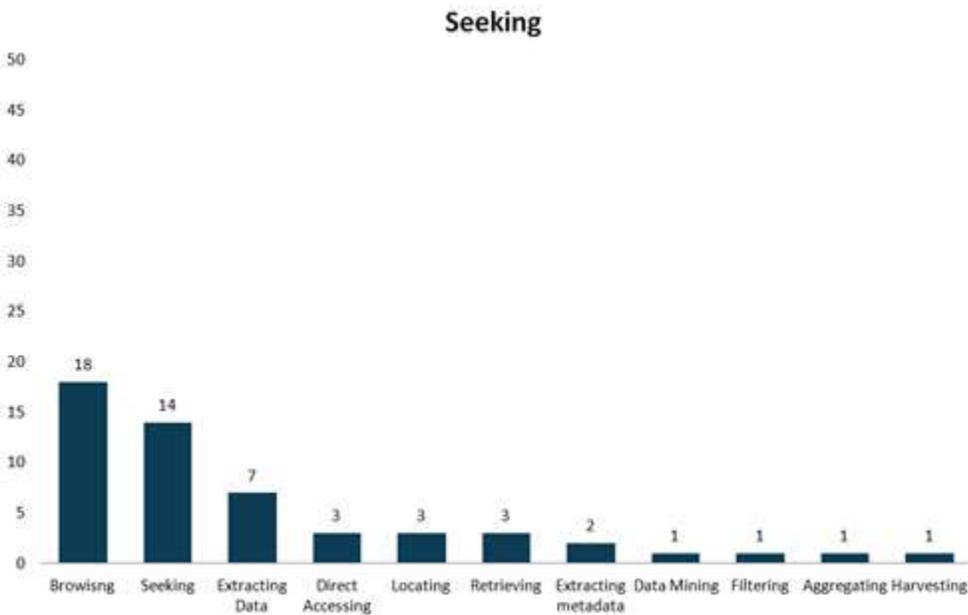

Figure 6. Distribution of the FLOSS tools according to the lower-level activities of the Activity Type "Seeking".

Overall, mapping the FLOSS tools registry against NeMO was an informative activity which enabled the enrichment of NeMO with new activity types, such as data extracting, metadata extracting, harvesting, aggregating and downloading. On the other hand, this alignment offers FLOSS, apart from an activity taxonomy for its tools, the potential of integrating those tools within a uniform conceptual framework for expressing knowledge about scholarly work, the NeMO ontology. The researchers' community will acknowledge the substantial contribution of the alignment as the final product would be a recommending tool, which would answer to the questions about the appropriateness of a tool accomplishing a specific activity.

4.1 Metadata field requirements

One feature of the FLOSS Inventory that was determined to need some additional work was the metadata fields. The original FLOSS Inventory contained 16 metadata fields and 5 categorization fields. Many of these fields including "Last Activity" and "Last Release" proved extremely time consuming to update regularly. This in turn meant much of the info on the Inventory was out of date. Furthermore, links to code repositories or documentation also were found to be out of date. Thus, after several discussions with developers within the EuropeanaTech community it was determined the that fewer metadata fields the better so long as the ones provided were kept up-to-date.

This section elaborates on our questions and conclusions on the topic of standardized metadata fields, necessary fields and quality.

4.2 Metadata fields

There is the need to evaluate available ontologies describing software tools (e.g. DOAP, dbpedia) and activities (e.g. DARIAH-NEMO, Premis … some examples of preservation activities can be found also here http://www.bac-lac.gc.ca/eng/about-us/preservation/Pages/preservation-activities.aspx)
Related to activities, we are not aware about having a controlled list of activities, using tags for this might be more appropriate, however there are taxonomies of roles in multimedia related applications … https://www.ebu.ch/metadata/ontologies/ebucore/index.html#Role

Important Metadata for getting the level of maturity of the tool are statistics like:
- Community size
- Number of know installations or companies using the tool

- User feedback in form of comments or popularity score

Other mandatory metadata should include (DOAP should be consulted for extending this list)

- Documentation: installation & user manuals
- License
- Source code repository
- Title
- Description
- Mime types
- Data formats

(the metadata from current Floss Inventory is already something that is recommended to be available for each tool)

5. Quality Control

It is not surprising that many tools developed during projects quickly enter the graveyard of Github after project funding ends. This leaves a large amount of promising software to age and become obsolete or be surpassed by others. Nothing is more disappointing than finding a description of the perfect tool only to find out that its last edit was made years ago. While it may still be of use, there are likely better options available. Furthermore, tools with poor documentation are of little value to most developers much like an item on Europeana with no metadata. How good is an open source tool if it's not clear what exactly it is able to do, and how can one use it?

That's why the FLOSS Inventory will only feature tools that are regularly updated and have proper documentation. To maintain a consistency when adding or removing certain tools the FLOSS Inventory will follow these simple guidelines:

Removal/Archiving
1. We will consider any tool that has not been updated in over a year to be inactive.
2. Before removing any tool we will reach out to the listed developer to inquire if the tool has been moved or is indeed inactive.
3. Any tool with insufficient documentation will be removed.
4. If a tool is found to have broken URLs and no new URLs are found the tool's owner(s) will be contacted and the tool will be removed until a functioning URL is provided.

Addition (tools must meet more than 1 of the requirements)
1. Tool must be relevant for one aspect of digital cultural heritage
2. Tool must have sufficient documentation
3. Tool must have been recently updated
4. Tools must be openly licensed

Lastly, as a general guideline we will favor tools and software that are popular in the network, trusted, and have a good community base.

Open Source Development is the norm for publically funded projects. Within these projects software is to be developed that will improve the standing not just for the developing institution but the sector at large. Millions of euros of funding go into the development of these open source tools and software. Disappointingly however, these tools usually have a poor impact, low developer uptake, and non-existent sustainability plans. This desperately

needs to change if Europe seeks to remain on the cutting edge of cultural heritage and digital humanities software development.

6. CMS functionality recommendations

For the past 4 years the EuropeanaTech FLOSS Inventory has resided in a Google Spreadsheet set up by the Netherlands Institute for Sound and Vision as part of Europeana v 2.0. While being a great place to organize data, a Google Spreadsheet is one of the least user-friendly platforms for any sort of browsing and searching. While we have spent the majority of the Task Force strengthening the data within the Inventory, it will all be rendered moot if there is no appropriate outlet to display it. It was requested that the developers within the Task Force write down recommendations for what they feel would make a good CMS for the FLOSS Inventory.

The following functionalities were considered to be particularly important for the FLOSS Inventory:
- Browsing content using a graphical user interface
- Sharing experiences with the tools
- Free text, tag- and metadata based search
- Tagging and user feedback
- Web API access
- System maintenance and sustainability

Browsing

The browsing functionality must offer multiple ways of exploring the repository.
- By tags
- NEMO Activity Types,
- Software categories (Pre-existing FLOSS Categories)

These all should be used for browsing the tool repository. The list of available tags and categories must be easily accessible and intuitively presented to the users (e.g. in form of tag clouds, or category tree).

Search

We suggest that different search and ordering alternatives need to be supported including:

- Free text
- Keyword search (tag/category)
- Metadata search (i.e. searching in a specific metadata field)
- Faceting
- Data/file formats (i.e. input & output)

Multifaceted search, is an important requirement, i.e. user can search or filter results by values of multiple characteristics of tools/services. For example, after searching for tools, which provide certain functionality, the user would likely next want to filter this by license. WordPress plugins can support this in combination with a DB, e.g.

https://facetwp.com/demo/cars/

Tags

It is very important to empower the CMS system with tagging functionality. While the categorization of tools using a predefined and fix hierarchical structure is generally useful, the tagging functionality is particularly useful for individuals and groups of users with very concrete interests. The tags are more dynamic concepts than categories, represent better the user's perspective and can be also used for different purposes (e.g. rating tools on a 5 level scale, marking what is the tool good for, grouping tools supporting a certain goal - i.e. the recipe names can be used for tagging tools-, etc.).

Searching tags and tools by tag labels is a very useful and common functionality in many CMS systems nowadays.

Experience sharing

In addition to being clear and concise the interface should allow users to simply browse large lists of tools and gain proper information,. It should also support users to share their experiences with the tools (e.g leaving comments, ratings). This may include recipes on how to combine different tools to achieve a certain goal, recommended plug-ins, wrappers, standardized interfaces, data conversions, etc. Recipes can be represented as sets of tools that, when combined, can do something greater and/or better than each component could separately achieve. For instance, IIIF[8], Blacklight[9], Mirador[10] and OpenSeadragon[11] are tools that can be used together or separately depending on desired functionalities.

---

[8] http://iiif.io/
[9] http://labs.europeana.eu/apps/blacklight-floss
[10] http://labs.europeana.eu/apps/mirador-floss
[11] http://labs.europeana.eu/apps/openseadragon

This is helpful because one of the major barriers to tool uptake is incompatibility. At least if input / output formats can be identified where relevant & also wrappers for different data models included in the list to help devs combine tools when needed.

API
The content of the FLOSS repository should be accessible also in machine-readable formats through standardized Web interfaces.
A (REST) API would be a useful asset, allowing client applications to access the FLOSS Inventory remotely in a well-defined way. The API should return JSON-LD format so that the same interface can be used both for integration in client applications and integration with linked data resources

7. Sustainability Recommendations

As with any platform it must be maintained, with respect to both content and installation. As a proposal, apart from the initial set-up and launch we suggest to plan:

- At least once annually, a content team checks and updates the information on all tools and services for the following 3 years.
- Hosting at an existing Web server, which is maintained by a Europeana partner and hence is maintained by an administrator, software is updated regularly, offline time is minimized.
- A SEO strategy for the platform (not sure how much this can be combined with any existing Europeana network marketing activity, e.g. social media?). Ideally there is a team, which is active for at least 6 months after platform launch with the task to increase visibility (and thus site visitors).

Also to share the wealth and information, we suggest submitting a tool review for publication in each issue of EuropeanaTech Insight[12], maybe even having a regular section in Insight about tools.

---

[12] http://pro.europeana.eu/get-involved/europeana-tech/europeanatech-insight

8. Additional recommendations

Throughout the Task Force many other ideas arose that would improve the FLOSS Inventory for both users and product owner. Those recommendations are listed below.

1. The original Task Force plan intended to provide reviews of the current listed tools. It was decided that doing this as part of the Task Force was not only not feasible due to time constraints, low personnel resources available in comparison to the size of the repository and no budget allocated for this activity. Also the members of the Task Force might not be the most appropriate persons to perform source code reviews and evaluate the quality of the software. We felt it better to leave this up to the community once there is an appropriate platform for the FLOSS Inventory.
2. There needs to be further evaluation of the available ontologies describing software tools like DOAP[13], Dbpedia[14] and activities such as DARIAH-NeMO, Premis[15], some examples of preservation activities can also be found on the Library and Archives Canada Preservation Activities website[16]. In relation to activities there needs to be more investigation into controlled lists.
3. Adding qualitative KPIs like e.g. user (i.e. developers) evaluations including trying to measure whether and how many developers within the GLAM community have found in the inventory what they have been looking for.

---

[13] https://github.com/edumbill/doap/wiki
[14] http://wiki.dbpedia.org/
[15] http://www.loc.gov/standards/premis/
[16] http://www.bac-lac.gc.ca/eng/about-us/preservation/Pages/preservation-activities.aspx

9. Future plans

In the future as part of EuropeanaTech there are several key plans for the FLOSS Inventory. One of the key aspects is community engagement activity. EuropeanaTech will coordinate with developer communities in the USA and EU like Hydra and IIIF. This will help EuropeanaTech better understand how to strengthen developer communities in Europe and better integrate the work being done around the world into the EuropeanaTech community. This will be done by participating in more workshops and events in addition to the continuation of Who's Using What columns and general editorial outreach.

EuropeanaTech will investigate conducting more user studies. What do developers need? How do they normally find tools? What would make things easier? How can we avoid parallel work and create a more collaborative developer culture within cultural heritage in the EU.

Further investigation of data models, structures and organization of the FLOSS Inventory.

Annex 1: Proto-personas

Added by: Sergiu Gordea

Persona	Application/Web Developer
Age	30
Skill Level (Years developing, knowledge)	5 years work experience
Sector (Museum, Library, creative, etc)	CH institution or SME reusing Europeana content
Desired outputs	API & Content

Added by: Sergiu Gordea

Persona	R&D employee (of CH institutions)
Age	40
Skill Level (Years developing, knowledge)	10+ development
Sector (Museum, Library, creative, etc)	(Large)National Libraries / Audio-Visual Archive
Desired outputs	(Metadata) Standards, Community support

Added by: Sergiu Gordea

Persona	Marketing Director
Age	35+
Skill Level (Years developing, knowledge)	5+ in market analysis
Sector (Museum, Library, creative, etc)	Museum
Desired outputs	Showcases & Demos

Added by: Agiatis Benardou

Persona	Digital Archaeologist
Age	35+
Skill Level (Years developing, knowledge)	10+ years experience in 3D visualisation, procedural and formal modelling to reconstruct and interpret ancient

	settlements.
Sector (Museum, Library, creative, etc)	Research / University
Desired outputs	

Added by: Agiatis Benardou

Persona	Researcher in Parliamentary Papers
Age	35+
Skill Level (Years developing, knowledge)	Basic knowledge of text-mining, annotation tools
Sector (Museum, Library, creative, etc)	LIbrary / University
Desired outputs	

Added by: Gregory Markus

Persona	Museum app developer
Age	28
Skill Level (Years developing, knowledge)	8
Sector (Museum, Library, creative, etc)	Museum
Desired outputs	Developer looking to find front-end CMS that will allow for neat and clean presentation of images and metadata from collection

Added by: Gregory Markus

Persona	AV Archivist
Age	35
Skill Level (Years developing, knowledge)	7
Sector (Museum, Library, creative, etc)	Audio visual archive
Desired outputs	Looking to find new OS ways to export collection

	metadata in RDF.

Added by: Gregory Markus

Persona	Student
Age	21
Skill Level (Years developing, knowledge)	1.5
Sector (Museum, Library, creative, etc)	Creative
Desired outputs	Student trying to find platform that will allow for live video annotations and way for the annotations to be exported and categorized for a research paper on crowd sourcing.

Added by: Zoltán Csáki

Persona	Head of department / Project manager
Age	35+
Skill Level (Years developing, knowledge)	10+ work experience, using pc and other digital devices and Internet regularly, has experience in dealing with and building databases
Sector (Museum, Library, creative, etc)	Library or Museum
Desired outputs	Institutional presence on the WEB (OPAC, Digital Library, connection to other services in the GLAM sector and the business sector), maintaining and developing institutional databases

Added by: Zoltán Csáki

Persona	Digital Librarian
Age	25+
Skill Level (Years developing, knowledge)	Degree in library science or humanities and information technology or familiar with information technology
Sector (Museum, Library, creative, etc)	Library or Museum or Software vendor

Desired outputs	Cleaned datasets, structured textual data (eg. TEI files), metadata mapping, optical character recognition, data or metadata conversion, bulk processing, schema definition

Added by: Marcin Werla

Persona	Employee of digitisation company
Age	27
Skill Level (Years developing, knowledge)	BSc in computer science, a bit of software development skills (simple PHP websites, some Linux scripting), working in digitisation project, responsible for improvement of company's digitsation workflow
Sector (Museum, Library, creative, etc)	Commercial company
Desired outputs	Improved scans of images, OCR

Added by: Marcin Werla

Persona	Curator in small (local) heritage institution
Age	40
Skill Level (Years developing, knowledge)	Art historian, engaged in promotion of history of his hometown, IT fan, recently set up his blog by installing wordpress on a virtual hosting facility (using precise step by step tutorial)
Sector (Museum, Library, creative, etc)	Museum/archive
Desired outputs	Something which will help him to promote his collections on-line - maybe a basic digital library system or some CMS popular in heritage sector.

Added by: Vladimir Alexiev

Persona	History educator
Age	40
Skill Level (Years developing, knowledge)	10 in specific domain, 1 in IT

Sector (Museum, Library, creative, etc)	Education (re Europeana for Education task force)
Desired outputs	Semantic search to find historic materials (eg caricatures, newspaper articles) by topic (eg the Eastern Front in WW1)

Added by: Vladimir Alexiev

Persona	Art researcher
Age	40
Skill Level (Years developing, knowledge)	10 in specific domain, 1 in IT
Sector (Museum, Library, creative, etc)	Visual arts (re future Europeana Art channel)
Desired outputs	Semantic search and classification to discover and classify objects related to art

Added by: Lyndon Nixon

Persona	Fashion magazine editor
Age	32
Skill Level (Years developing, knowledge)	10+ in fashion, low technical knowledge
Sector (Museum, Library, creative, etc)	creative
Desired outputs	Illustrative image search for fashion (trends)

Added by: Lyndon Nixon

Persona	(AV) Content producer
Age	47
Skill Level (Years developing, knowledge)	15+ in TV/media
Sector (Museum, Library, creative, etc)	creative
Desired outputs	Documentary or news creator with a need for easy (re-use) access to historical media

Annex 2: Original FLOSS Scoring Guidelines

Quality of documentation
- Hosting and find ability (Github, custom website), e.g. http://annotorious.github.io/
- do the docs contain the most necessary information?
- are the docs well written?
- Frequency of documentation updates
- Tutorials for users
- Tutorials for programmers, code examples e.g
https://github.com/okfn/annotator/wiki/Getting-Started
- Live demo, e.g. http://annotorious.github.io/demos.html

Ease of adaptation
- Ability to plug-in (custom modules), e.g https://github.com/okfn/annotator/wiki/Plugin-Development
- Ability to be plugged on other tools (e.g API) e.g.,
https://github.com/okfn/annotator/wiki/Storage- Mailing list / forum
- Documentation of for adapting/extending
- Dev support e.g. via mailing list https://groups.google.com/forum/#!forum/annotorious
- Number or list of known installations, with focus on production/commercial uses.

Project setup quality:
- Code Hosting (e.g. Github)
- is the package structure clear?
- are there any comments in the code? (if not, is the code readable?)
- Programming Language (common vs. exotic)
- Existence, and other metrics (such as code coverage) of automated tests
- does have convenient build scripts? (is there anything that indicates it's easy to run?)
- Use of standards (e.g. for data exchange)
- Number of contributors
- Frequency of code updates/releases, e.g. https://github.com/okfn/annotator/releases
- Existence of active community, e.g. for issues:
https://github.com/annotorious/annotorious/issues
- Sustainability plans / existence of a product roadmap