Digital Preservation Activities and Workflows

EMBL Archive, 2021

Table of Contents

1. Document Statement of Purpose

The purpose of this document is to record the specific procedures, workflows and standards in place in May 2021 relating to born-digital archives to ensure continued access to this material. It is designed to complement the more general Activities and Workflows for the EMBL Archive. It also provides some guidance on the implementation of the EMBL Archive Terms of Reference and the Digital Preservation Framework.

1.1. Digital Preservation

Digital preservation is defined as a series of activities that seek to ensure continued access to digital assets. It includes the policies, workflows, people and systems involved in endeavouring to make sure that digital assets are accessible and reliable in the long term. Digital preservation aims to ensure that information that exists in digital format remains accessible, usable and reliable in the medium and long terms. Within the EMBL Archive, this activity is guided by the EMBL Archive Digital Preservation Framework, which outlines the strategic principles of digital preservation in the EMBL Archive and details responsibilities. In addition, the framework also provides details on the two models that guide this activity in the EMBL Archive, namely:

- The Open Archival Information Systems Reference Model (ISO 14721:2012 – Space data and information transfer systems – Open archival information system (OAIS) – Reference model), which describes different components of a system responsible for long-term digital preservation.
- The Digital Curation Centre's Digital Curation Lifecycle Model, which presents sequentially the activities that are needed in the process of preserving digital assets.

1.2. Organisation of this Document

This document is divided into three sections, according to the three objectives of the EMBL Archive, which are to capture EMBL's documentary heritage, to manage EMBL's archival holdings and to make these holding accessible. Each section includes a list of the activities under each objective, and then a description of the relevant procedure, workflows and standards is provided.

2. Manage EMBL's Digital Archival Holdings

The activities undertaken by the EMBL Archive, relating specifically to born-digital archives, are:

- Documenting Digital Accessions
- Describing Digital Archival Holdings

- Archival Appraisal of Digital Materials
- Logging Physical Storage
- Deaccession
- Disaster Preparedness
- Overseeing IT Systems
- Security Matters
- Managing Personal and Sensitive Data

2.1. Documenting Digital Accessions

Accessions are materials (analogue or digital) which, following their transfer to the EMBL Archive, become the physical and intellectual responsibility of the EMBL Archive. The accession is marked by the creation of an accession record. From 2020, accessions are documented using the accessions module of the EMBL Archive instance of the AtoM application (archive.embl.org). The information entered in the accession record can then be used as the basis of that material's archival description and the accession record and the archival description can be linked.

In addition to creating an accession record, further information relating to each accession is added to the Planning Matrix (see: Archive Drive > Processing Documentation > PlanningMatrix). Whilst this appears to be a duplication of effort capturing details in the planning matrix allows for additional considerations to be noted including any potential sensitive content, the need to appraise the content and any hardware or software issues that may have arisen during the accession process. This helps to ensure that visibility of the born-digital archives remains high and identifies the next processing steps for the content. If additional hardware is needed to read the media or the accession includes a file format not encountered before it will be noted in the Planning Matrix – having this information to hand is crucial for planning purposes. The spreadsheet format also allows for filtering of information in a more flexible way than the accession module does.

When digital material is accessioned, it is read on the digital forensic workstation. If appropriate, the write blocker will be used to preserve the accession's integrity. A digital copy of the content will then be created and stored on the dedicated USB stick. From this copy and as part of the accession of digital material, the following software will be used to create reports:

- DROID is a software tool developed by the UK National Archives and released on an open-source basis. It is a file format identification tool that identifies the file type and version from information embedded within the file itself rather than relying on the file suffix. The software can be used to create a file manifest of the collection prior to any processing or ingest. It will also create a checksum for

each file which can be used to check that the file contents remains unchanged over time – changing a single character in a 200 page report would generate a different checksum. DROID should be used as soon as is practical after the decision to accession the material has been made. Running the tool creates a list of files including the identification of the file format, details of which can be compiled in the PUID analysis spreadsheet (see: Archive Drive > Processing Documentation > DigitalHoldingsReportsAndPUIDanalysis > PUIDanalysis_v3. Noting the number of files for each identified file format provides the EMBL Archive with some crucial perspectives of the complete collection of born-digital material. Being able to identify the most common file formats held and which formats appear most frequently across the accessions allows the archivist to make planning and prioritisation decisions. For example, the archivist will be aware that it has received some email content but may decide to wait until it has a critical mass, measured either in quantity of files or frequency of occurrence before investigating appropriate tools for the preservation of this format.

- Archifiltre is an open-source tool that supports the appraisal of digital archives by creating file trees, flagging duplicates and allowing the reorganisation of directories. Its visual representation of folder sizes and file types is especially helpful when faced with a large quantity of material to review. Using the tool, especially for large or complex born-digital accessions will be useful to make an initial assessment relating to the appraisal of the collection. Its visual interface helps the archivist to browse the accession and allows for the identification of directories that might contain a large number of files or very large files. Observations gleaned from the Archifiltre analysis are noted in the planning matrix. A .json file can be created and saved for future use.

From the digital forensics workstation, the reports created above are to be transferred to Archive Drive > EMBL Archival Holdings > Processing Documentation > DigitalHoldingsReportsAndPUIDanalysis > [Directory Name = Accession Number] > [Accession Number]_Archifiltre / DROID.

2.2. Preparing Digital Material for Ingest

Preparing digital holdings for means ingest taking the digital holdings analyses and creating an original version that will be ingested by Archivematica. Archivematica is, in essence, the safest place to hold digital material. Though in principle, what is transferred and ingested into Archivematica should be already appraised, if it is not possible to do so in a realistic timeframe, unappraised material should be ingested into Archivematica for safekeeping – to be appraised at a later date. When this is case, it will not be necessary to normalize the material at this stage, given that the normalization process is preparation for long-term preservation and / or access to DIPs. (See 2.7. Archivematica and AtoM.)

Normalization choices are managed with Archivematica. Where Archivematica does not have a pre-approved normalisation path for a given format, a workaround is to create an access copy prior to ingest and include this copy in the transfer package. The current processing configuration has been set so that a decision needs to be made manually regarding normalization during ingest.

Appraisal of digital material is intellectually similar to that of analogue material, but technically different. The Archifiltre and DROID reports will help identify duplicates and directory structures that can be used for cataloguing. In particular, file formats such as DS_store file (MAC) (fmt/394) and Thumbs DB (fmt/682), which are both outline records and not worthy of long-term preservation, can be deleted at this stage.

Appraisal follows the principles outlined in the EMBL Archive Terms of Reference and the Activities and Workflows document. Throughout the appraisal, it is important to keep a log of what is deleted in order to record this as part of the archival description. The file PUIDanalysis_v3 should be updated when material is deleted to reflect the true nature of the digital holdings.

It is to be expected that some digital files will not be accessible (e.g. cannot be opened, file already corrupted, hardware cannot be read). In such cases, the EMBL Archivist may need to undertake further research or spend extra funds in order to make an informed judgment call on how to proceed.

The copy of the accession can be used either to "draft" the appraisal of material, or can become the new master copy that will be ingested. As good practice, it would also be appropriate to run new Archifiltre and DROID analyses, generate reports and store these in addition to the first reports. This is because the original reports that were created at the time of accession will no longer be applicable to the material that is actually ingested, preserved and made available in the long term.

This process assumes there is an endpoint in AtoM to link to ingested material. At present, this is only possible at collection / fonds level. If DIPs are to be visible in AtoM, make sure there is an appropriate archival description to link it to before starting the ingest. (The collection slug is required for input during ingest.)

Note: Archivematica cannot directly access the Archive drive. Therefore, it is necessary to copy the directories that will be ingested to a different drive using software such as Cyberduck. Only once this is done will it be possible to select files at the start of the transfer process in Archivematica when "Browse" is clicked.

2.3. Describing Digital Archival Holdings

This activity builds on 2.1. Documenting Digital Accessions and is linked to the 3.2. Metadata and Taxonomies and 2.10. Managing Personal and Sensitive Data activities described below.

The purpose of this activity is to create archival descriptions, catalogued in the EMBL Archive catalogue that facilitate discovery of and access to the holdings of the EMBL Archive and provide information about their availability for consultation.

Archival descriptions are created and maintained in the EMBL Archive instance of AtoM (archive.embl.org). The cataloguing standard used by the EMBL Archive is the General International Standard Archival Description (ISAD(G)) maintained by the International Council on Archives. When appropriate, the ICA's complementary standards are also used. These include the International Standard

for Describing Functions (ISDF), the International Standard Archival Authority Record for Corporate Bodies, Persons and Families (ISAAR (CPF)) and the International Standard for Describing Institutions with Archival Holdings (ISDIAH).

By their nature, born-digital archives are likely to require different level of cataloguing than has been the case with analogue holdings in the EMBL Archive. The sheer quantity of born-digital material means it might be appropriate to adopt different approaches, for example:

a. Faced with a large number of related photographs it might be sufficient to describe the material at series level describing the material in its broadest sense. If there are no copyright issues and the material will be available for viewing online, it might not be necessary to describe each item if a user is able to access the original content directly. Many archives already adopt a similar approach for example providing a single description for a bundle of letters.
b. In some cases the digital content may be a digital version of an item that is already held in analogue form by the archive. This means that the existing entry will require only a minor modification to include reference to the digital asset.
c. Some digital content may require more work to provide a suitable description. For example, with an oral history recording a brief summary of an hour-long discussion may not be sufficient for a researcher to determine its relevance. It may be that a full transcript, whether generated manually or through the use of voice recognition software is required. If the transcript becomes part of the index record then there is a much higher likelihood that it will feature in search results and you would expect use to increase.

It is only to be expected that practice will evolve with regard to the cataloguing of born-digital material held in the EMBL Archive. As with analogue materials, the approach and the time you might allocate to a collection should be expected to vary from collection to collection depending upon the nature of the material and its relative significance to the organisation. As more born-digital content is described there will be greater awareness of what level of cataloguing is appropriate. This is also an area to discuss with users in terms of their discovery and use of born-digital material.

2.4. Archival Appraisal of Digital Materials

Archival appraisal is carried out after accessioning, during the process of archival description (see: 2.2. Describing Archival Holdings).

The purpose of appraisal is to retain material that is aligned with the scope of the EMBL Archive, as described in the EMBL Archive Terms of Reference, and dispose of material that is out of scope. When this is the case, information about the disposal is recorded directly within the archival description of the holdings at fonds or collection level as a bare minimum.

Appraisal decisions are based on the content of the material and not its physical format but there are some specific issues arising from the appraisal of born-digital archives. These include the ability to scan thousands of files to detect exact duplicates (via the file's checksum) allowing the Archivist to decide whether these files should be removed from the collection before the content is ingested into Archivematica. There will also be a necessity for work to be undertaken on some media and with some file formats to enable an informed decision to be made.

A timely review of material as soon as it has been accessioned into the EMBL Archives will highlight the existence of media or file format issues. If an issue is detected there might be an opportunity to see whether it is possible to secure the content in an alternative format to allow the appraisal to be undertaken but it is recognised that this will not always be possible.

Some institutions have a policy of taking photographs of floppy disk drives where the label is not blank. Whilst recognising that the label and the contents of the disk may not necessarily be the same such a policy isn't too dissimilar from writing details from an arch lever file before removing the contents from the file. If the decision to adopt such a policy at EMBL the process can be quickened through the use of a template printed onto acetate and then adding the reference number for the media using a non-permanent marker that can be easily wiped clean with a damp cloth.

Where appraisal is undertaken prior to ingest into the Archivematica instance, it is good practice to create another DROID report and create a second file manifest to capture details of the content that is actually ingested. In such situations the PUID analysis spreadsheet should also be updated so it reflects the collection as held.

2.5. Logging Physical Storage

A key principle when dealing with born-digital archives is the policy to remove content from their original media to a more stable format. To support this, the EMBL Archive has installed a forensic workstation, a dedicated PC that is used solely for processing born-digital archives. A range of software, mostly open-source, can be used to support processing tasks. Researchers will also be looking to use software and tools to gain new perspectives into the material in archival care. One essential characteristic of a forensic workstation is that it off-line. With a dedicated hard drive for the storage of transferred digital archival holdings. From this drive, the material can easily be copied onto an external drive and stored at a different location, it will later be ingested into the Archivematica instance.

Digital material needs to be stored safely with restricted access (Room 14A-005) prior to the content being ingested into *Archivematica*. In cases where digital material is included amongst a paper file then the use of an insertion sheet will ensure that users are aware of the existence of digital material in the paper file.

Even after removing the content from removeable media the policy of most institutions is to retain the original as a digital artefact and they are often used as part of its digital preservation advocacy work.

2.6. Deaccession

To deaccession material is the process by which an item which has been accessioned into the EMBL Archive is permanently removed from the holdings either relocated to another Archive or returned to the depositor. This process requires the approval of the EMBL Director General and the EMBL Archivist. The deaccession is documented in a form (*EMBL Archive Deaccession Form*) which details the reason and the date of the deaccession.

Analogue material can be physically returned to a depositor or transferred to a more appropriate archive service with appropriate permissions. With digital material the ease by which the content can be copied makes deaccessioning much harder to demonstrate. The remaining steps are dependent upon the degree to which the material has been processed:

- If the material has not been catalogued or made available, the process is quite straight forward. The entry in AtoM would need to be updated to reflect the change in status. In addition, the archives should consider retaining the DROID report of the accession as evidence of what had been held temporarily. All files should then be returned to the depositor or deleted as agreed.
- If the material has been catalogued and crucially made available, the situation is a more complex. In addition to updating AtoM there is a need to consider the impact of removing the links and the likely disruption and confusion this will have with users. With born-digital content published online has a persistent identifier the archive will look to remove the material but leave a 'tombstone page' including a statement of unavailability that details the reason for the item's removal.

2.7. Archivematica and AtoM

Archivematica is enabled to display DIPs (access copy derived from the ingested file) in AtoM. For this, Archivematica has its own user login and password for the AtoM instance.

Displaying ingested material in the EMBL Archive catalogue is made possible by linking the (Archivematica) DIP to the collection slug (in AtoM) either at the very start of the transfer ("Access system ID") or later during ingest ("Upload DIP"). When an AtoM slug is provided, Archivematica correctly creates new entries for each digital object. From observation, these are always at item level, and hierarchies are not automatically inherited.

Following this, it is necessary to manually organise the objects in order to link them to existing or new archival descriptions. These objects will need to be treated differently depending on whether it is digitised material or if it is born-digital. Indeed, in AtoM, it is possible to move a digital object into an existing hierarchy (archival description). This correctly creates a second entry so that there is now one for the analogue item and one for the digitized image. This is potentially confusing to the user, so archival descriptions need to include information about under the relevant "Access" tab.

If the ingested objects are born-digital and there is no pre-existing archival description, the archival description created through the DIP upload become the only archival description for these objects. Descriptive metadata needs to be added to complete the catalogue entry and facilitate access.

At the moment, there is no function within AtoM to merge the technical-metadata Archivematica entries into AtoM's descriptive metadata entries even if they refer to the same object. Therefore, this most of the work has to be done manually with the "Move" feature. On the other hand, this offers flexibility to create custom-made access solutions for each ingest. See: 3.2 Metadata and Taxonomies and the Archivematica documentation.

2.8. Disaster Preparedness

The EMBL Archive is a founding member of NARN, the Notfallverbund Archive Rhein-Neckar e.V. (Emergency Network for Archives in the Rhine-Neckar Region, https://narn.eu). The mission of NARN is the protection and salvage of cultural assets in the event of a disaster impacting any member institution. The activities of the group include knowledge exchange, drills, liaising with the fire service, purchasing insurance services and sharing salvage equipment (stored in the Heidelberg Stadtarchiv). As a registered association, NARN has statutes (*Satzung des Notfallverbundes Archive Rhein-Neckar*), agreed to by all the founding members in 2018.

In case emergency access to the EMBL Archive storage room is needed, a key to B14A-005 (the storage room) is kept at EMBL Reception.

Due to its format, born-digital archives are in a different position to analogue archives when it comes to disaster preparedness. Whilst the material exists only on removeable media it is as much at risk as its paper equivalents. Once the content has been ingested into Archivematica the content will be stored in the AIP-store and it falls under the responsibility of EMBL IT Services with this network location being backed-up according to its own practices and policies.

2.8.1. In Case EMBL (or Its Successor Institute) Ceases to Exist

The EMBL Archive's holdings bear witness to one of humanity's greatest adventures: the scientific study of the natural world. In case EMBL, or any institute that might come to replace it, ceases to exist, these holdings should be transferred to an adequate place for long-term preservation and access. In March 2020, the EMBL Archive and EMBL Legal Services started exploring ideas of how to best address this. This paper will be updated in due course to document any decisions made in this regard. Where content is transferred under these circumstances the creation of a tombstone record (see 2.5. Deaccession above) to reflect the changes is not likely to be a feasible option.

2.9. Overseeing IT Systems

The EMBL Archive uses three specialised software to deliver its missions:

- AtoM (https://www.accesstomemory.org/en/) is an application for archival description. The EMBL Archive instance is accessible at archive.embl.org and is the EMBL Archive Catalogue.
- Archivematica (https://www.archivematica.org/en/) is an application for digital preservation. The EMBL Archive instance is accessible internally only at digipres.embl.org.
- The Archivematica Storage Service, which is used by Archivematica to manage archival storage, transfer locations and packages. The EMBL Archive instance is accessible internally only at https://digipres.embl.org:8000.

All three software are hosted and maintained at EMBL by IT Services, but the EMBL Archivist has administrative rights for all of them. It is the responsibility of the EMBL Archivist to manage these IT systems as much as possible and liaise with IT Services as and when needed.

2.10. Security Matters

To ensure its trustworthiness and reliability of the EMBL Archive, its holdings must be kept safe and secure. To this end, measures are taken to limit access to the digital holdings with access to the EMBL Archive server (smb://ad-fs01.admin.embl.de/Archive) password protected to ensure that only authorised individuals can access it. This server is managed and backed up by EMBL IT Services.

The security details of Artefactual products are detailed as part of the online documentation

- Archivematica: https://www.archivematica.org/en/docs/archivematica-1.12/admin-manual/security/security/
- AtoM: https://www.accesstomemory.org/en/docs/2.6/admin-manual/security/

2.11. Managing Personal and Sensitive Data

Due to their nature, archival holdings often contain personal and sensitive data (such as dates of birth, private correspondence or evidence of personal opinions) and the EMBL Archive is no different. To ensure that this type of data is well managed and that the identity of individuals and sensitive data about the institution are protected, the EMBL Archivist works with the EMBL Data Protection Officer to ensure that robust procedures are in place and that these are documented. To this effect, the EMBL Archive prepared in 2018 the *EMBL Archive - Record of Processing Activities*, in pursuance of EMBL Internal Policy 68 on General Data Protection. This document was published online in 2019.

When access to archival holdings is restricted because they contain personal or sensitive data, this is indicated in the EMBL Archive catalogue.

The management of personal and sensitive data within an archive is an activity that needs to be constantly monitored and regularly reviewed. Therefore, the EMBL Archivist and the EMBL Data Protection Officer meet on a regular basis to discuss any matters arising. The sheer quantity of born-digital archives and the fact that the information is more likely to be recent does increase the likelihood of it including both personal and sensitive data.

Consideration for personal and sensitive data needs to form part of the initial review and assessment of the collection. In cases where a collection does (or is likely to) contain sensitive content it is important to determine whether this relates to a specific section of the collection or across all of the material. In cases where the sensitive content is contained within a section it may be possible to effectively split the collection into two and apply different methodologies to the two sections.

Where paper archives containing sensitive content, emphasis is placed on the physical environment to protect the material until its sensitivity has passed. However, with born-digital material there is a need to actively manage the content during its life even if access is not likely for twenty years or more. One potential solution might be to ingest this content into Archivematica but to not complete the processing, so it has an AIP bit not DIP. The content will then be backed-up by virtue of being in the AIP store and can be managed with regard to file formats without a risk of it being accessed prematurely.

3. Make EMBL's Digtial Holdings Accessible

The specific activities undertaken by the EMBL Archive to deliver this objective are:

- The EMBL Archive Reading Room
- Metadata and Taxonomies
- Digitisation
- Communication and Outreach Activities

The Digital Library Federation's *Levels of Born-Digital Access*, published in 2020, has quickly been adopted by the digital preservation community. It provides a framework of five areas; Accessibility, Description, Researcher Support & Discovery, Security and Tools across three levels Level 1 (the lowest level of compliance) to 3 (the highest level of compliance). The guidelines explicitly state that Level 3 "will not be appropriate for all institutions."

In adopting this framework, the EMBL Archive aspires to achieve Level 2 across all five areas:

Area	Description	Notes (to get to this level...)
Accessibility	Researchers are provided with information on the accessibility of born-digital materials and able to find accessible information and metadata about the materials they need.	Website compliance to accessibility standards? Ease by which screen readers, colour contrast can be applied to the online catalogue pages, adding alternate text to embedded images, or adding tags to define reading order. What does accessibility mean for Oral history material (for example) – [for example transcripts = Level 3] Review which field(s) in AtoM to use to support this work.
Description	Provide required descriptive elements for a collection-level record and at least one descriptive note about the processed digital materials and a multilevel record and specified descriptive notes	Suggests adding text to *Physical Access, Technical Requirements* and *Conditions Governing Access/Use* fields to describe how the materials can be accessed. If the materials are inaccessible because they are unprocessed, explain how users can request access (e.g., send email to department@email.com). Any copyright issues if you provide access without knowing who is looking at the material?
Researcher Support & Discovery	Support basic access to and duplication of content. Have a knowledgeable staff member available to provide assistance. Create and provide access to additional technical documentation. Have multiple staff available to provide assistance. Seek feedback from researchers.	Information about accessing and using born-digital archives on archives website (re-use for searchroom visitors) All staff supervising the searchroom to be aware of any specific issues relating to accessing born-digital archive content Decide on policy about copying items/entire born-digital collections and ensure this is part of the general information on the archives website Identify software for accessing content [I would suggest this is only for content not in common formats mp3, jpg, PDF] so may just be a flag to say check format(s) as part of making born-digital content live? Feedback might be formal (surveys) or informal and might relate to process and the experience. As the service is being developed is an excellent opportunity to engage with users.
Security	Provide access to open, authentic, virus-free content on a dedicated on-site public access PC with security measures implemented based on local policies	Virus scan the access versions of the born-digital content? Publish checksum of the access file so a user can verify the file is authentic? There is an on-site public access PC which requires non-admin login for use in patron access

Area	Description	Notes (to get to this level...)
	Provide secure remote access to open content using platforms that meet local requirements for secure remote access based on institutional policies and access control needs	Is browser cache automatically cleaned-out with session log-off / log-on Is USB drive accessible on public access PC?
Tools	Provide local access via an on-site public access computer with open and common software to render widely used file formats. Provide tools to support enhanced access and exploration of content, and software to support evaluation and investigation of files; provide mechanism for semi-mediated distribution.	Provide open and common software to render widely-used file formats on local PC Provide applications that support access to a wider variety of file formats, specifically considering formats that may be locally specific. Provide software to support enhanced evaluation and investigation of materials. Provide mechanism for simple distribution of materials, such as email and cloud-based file sharing services. Need to decide if Archives are to provide access to DROID export, disk image and other underlying metadata

3.1. The EMBL Archive Reading Room

The EMBL Archive reading room (B14A-004) is open to any member of the public who wishes to consult archival material. Appointments should be made ahead of time. Persons consulting material will always be supervised. No food or drink is allowed in the EMBL Archive reading room.

It is anticipated that most engagement by users of the born-digital content will be via the website. A desktop computer is available for public use in the reading room: an Archive Guest log-on and password can be obtained from the EMBL Archivist.

3.2. Metadata and Taxonomies

Archival metadata helps describe archival holdings, making them discoverable and accessible. All metadata and taxonomies used in the EMBL Archive are managed via its AtoM instance.

The AtoM application can generate Dublin Core 1.1 XML and EAD 2002 XML for all archival descriptions, regardless of their level. The application also allows for the management of taxonomies. The two taxonomies that are currently used classify:

- Places (which appear as “EMBL sites” on the menu): Each EMBL site has a term in this taxonomy.
- Subjects: This is for general areas relevant to the EMBL Archive. Each Subject term has a scope note.

In addition, seven functions are defined and have a scope note. Finally, an authority record is created for the creators of material transferred to the EMBL Archive.

Archivematica uses the Metadata Encoding & Transmission Standard (METS), to act as a wrapper around other metadata content including Dublin Core and PREMIS. PREMIS is the international metadata standard to support the preservation of digital objects and records actions (called Events) undertaken on the digital objects by Agents (which includes software) and Rights (including copyright and licenses).

It is possible to add descriptive and rights metadata during the ingest process but this is applied to all digital objects included in the transfer. Applying such metadata post-ingest via AtoM might be more appropriate. See: 2.7 Archivematica and AtoM and the Archivematica documentation.

3.3. Digitisation

Although the primary focus of this paper is on digital preservation, Archivematica is designed to work with all digital content, regardless of if it is born-digital or digitized. It makes sense for digitised content to use the system so there is a single route through which digital material is ingested and made available to users via the AtoM discovery interface. It is possible to bypass some stages of the ingest process, so if you have access versions for the digital objects these can be included in the transfer if the reserved folder name conventions are followed:

Accession top folder
- objects [folder]
- Access [folder]
  - photo1.jpg
- photo1.tif

In this example an access copy of photo1.tif already exists, by including it in the transfer package Archivematica will use photo1.jpg and will not create a new access version as part of the ingest and normalisation workflow.

3.4. Communication and Outreach Activities

The EMBL Archivist works with EMBL Communications to help ensure that news about the EMBL Archive is shared with the EMBL community, the archival community and the wider world. News stories about the EMBL Archive are collated on the EMBL News website with the tag ‘EMBL Archive’ and can be accessed through this link: https://www.embl.org/news/tag/embl-archive/.

In addition to news stories, the EMBL Archivist reaches out to the archival community and professional associations through mailing lists and social media, as and when relevant. On Twitter, the hashtag #EMBLArchive is used to reference the EMBL Archive, and all tweets can be accessed here: https://twitter.com/hashtag/emblarchive.

There are a number of internal and external advocacy opportunities surrounding the provision of access to born-digital materials. As the Archives experience increases then further information can be placed on the archive webpages about the use of the material. Consultation with users is also an explicit element of the DLF's *Levels of Born-Digital Access*.

4. Further Activities

4.1. Institutional Committees

Digital preservation is a challenge faced by the whole organisation. There is an opportunity for the EMBL Archives to undertake advocacy about the issues and the opportunities associated with digital preservation including preserving content for future generations and the new perspectives that applying tools on historical data.

4.2. Associations and Collaborations

As in many institutes, the EMBL Archive maintain connections with colleagues and organisations outside of EMBL. This ensures staff keep their professional development up to date through knowledge exchange, professional practice and standards. Membership to professional organisations also means agreement to the association's professional code of ethics. Importantly, the EMBL Archivist, as a member of the International Council on Archives, is bound by its Code of Ethics (available at: https://www.ica.org/en/ica-code-ethics) It is important to note here that ICA, in its activities and statements, does not differentiate between analogue or digital records and archives – digital material are part of the humanity's heritage and needs to be preserved ad protected appropriately.

4.3. Digital Preservation Coalition

The Digital Preservation Coalition (DPC) develops and maintains resources like the DPC Handbook and technology watch reports, with regular blogs by members sharing practical experiences. It organises many events, most of which are for DPC members only, but the slides are usually available for members and non-members alike. In the future, EMBL might consider joining the Digital Preservation Coalition to develop its own digital preservation activities and to contribute to the development of the discipline.

5. Revision of this Document

This document was prepared between April and June 2021 by the EMBL Archivist and should be reviewed annually.

6. Appendix

Appendix 1: Summary flowcharts for the principal workflows in the EMBL Archive