

PARALLEL ARCHIVE

beta

Gabriella Ivacs, OSA, ivacs@ceu.hu

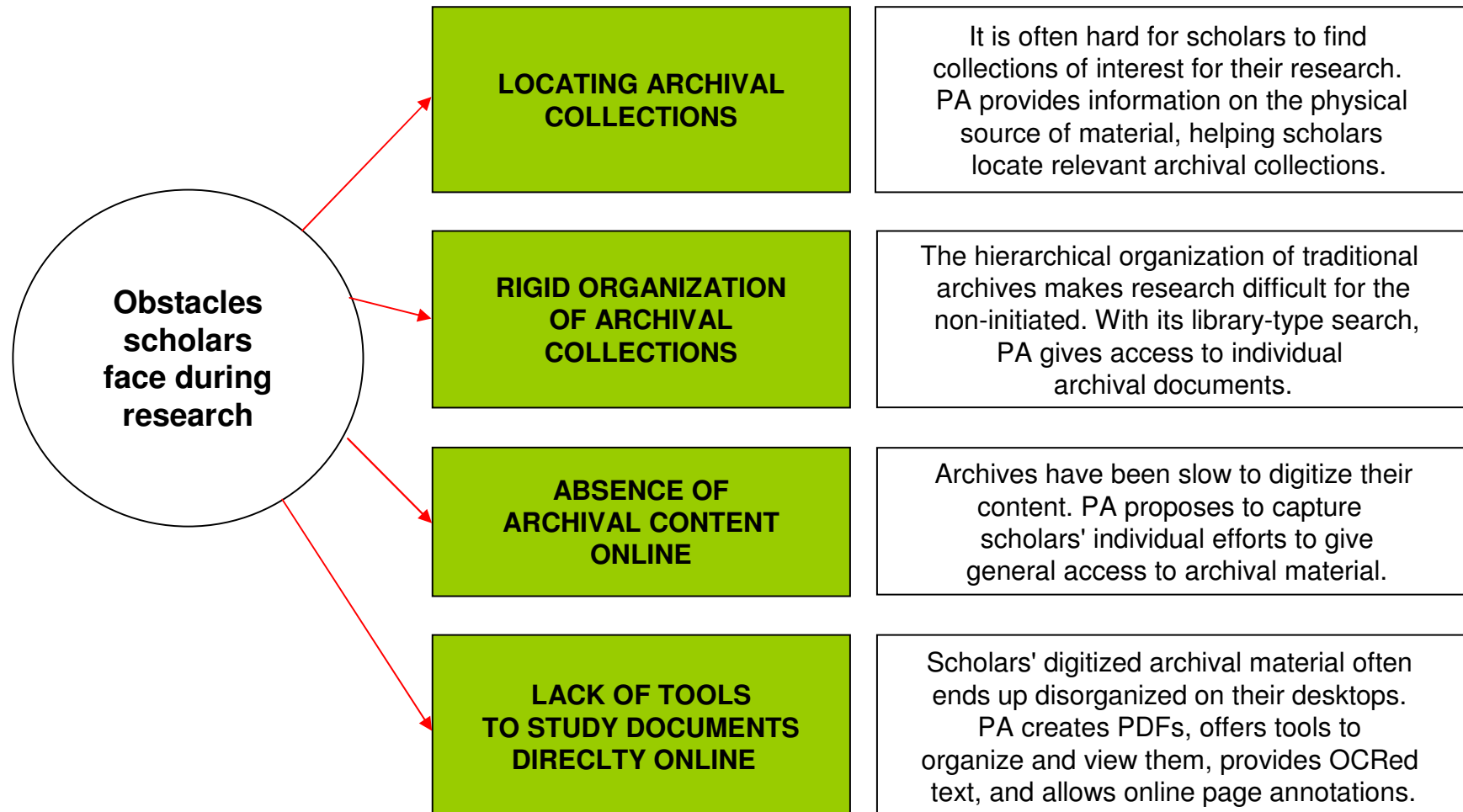
Sofia, September 16, 2009

The PARALLEL ARCHIVE project: www.parallelarchive.org

- Developed by the Open Society Archives@CEU, funded by the Institute of Record, maintained, powered by an external company and CEU
- OSA Laboratory Project, it is not the OSA Digital Repository
- 2007: consultation with archival institutions, scholars, fundraising experts, developers
- 2008: specifications, project planning, implementation, user testing
- 2009: sustainability planning, 3-year roadmap, fundraising, prototype roll-out at small research communities
- 2010: dissemination plan, sustainability plan, governance model and development (batch ingest, more research tools)

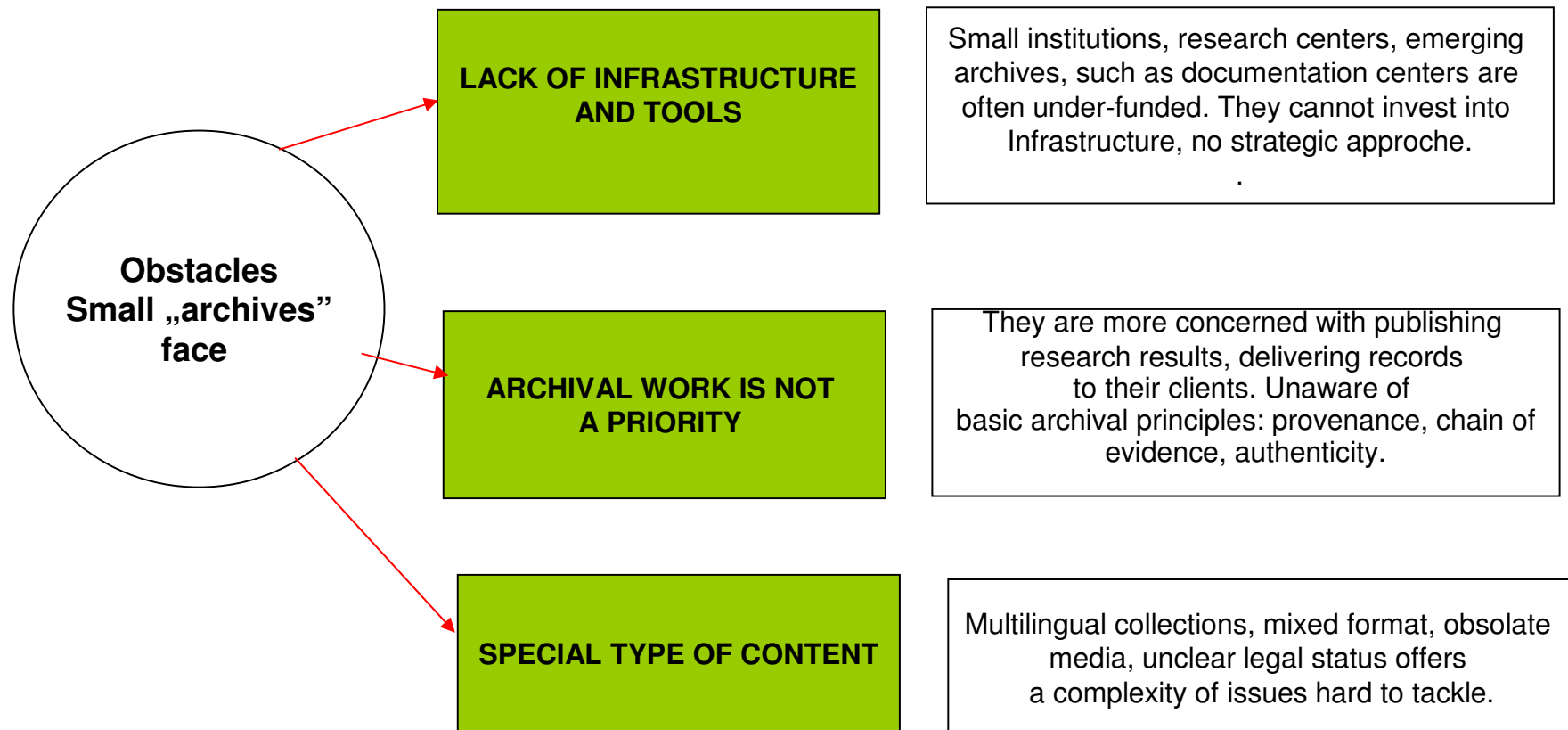
WHY DO WE NEED A PARALLEL ARCHIVE ?

PA was created to address the problems scholars face when working with archival material.



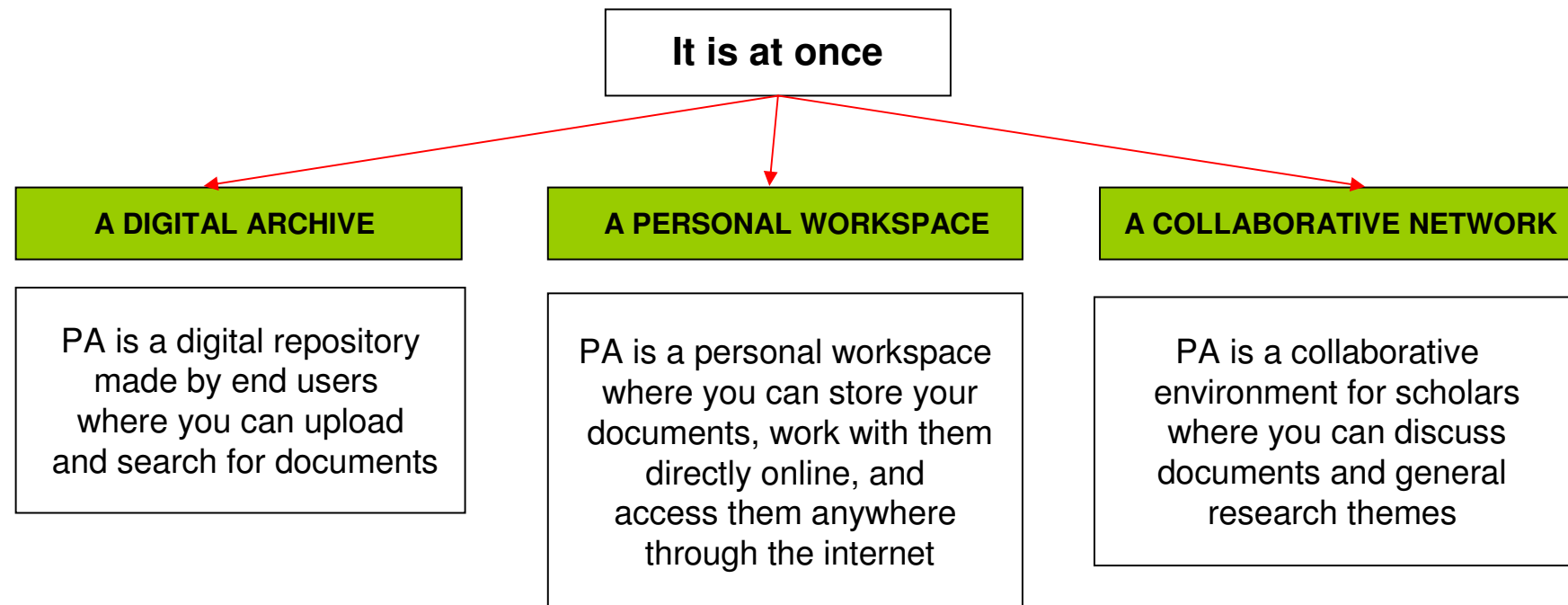
WHY DO WE NEED A PARALLEL ARCHIVE ?

PA was also created to address the problem of „small” collections, archives, research institutes

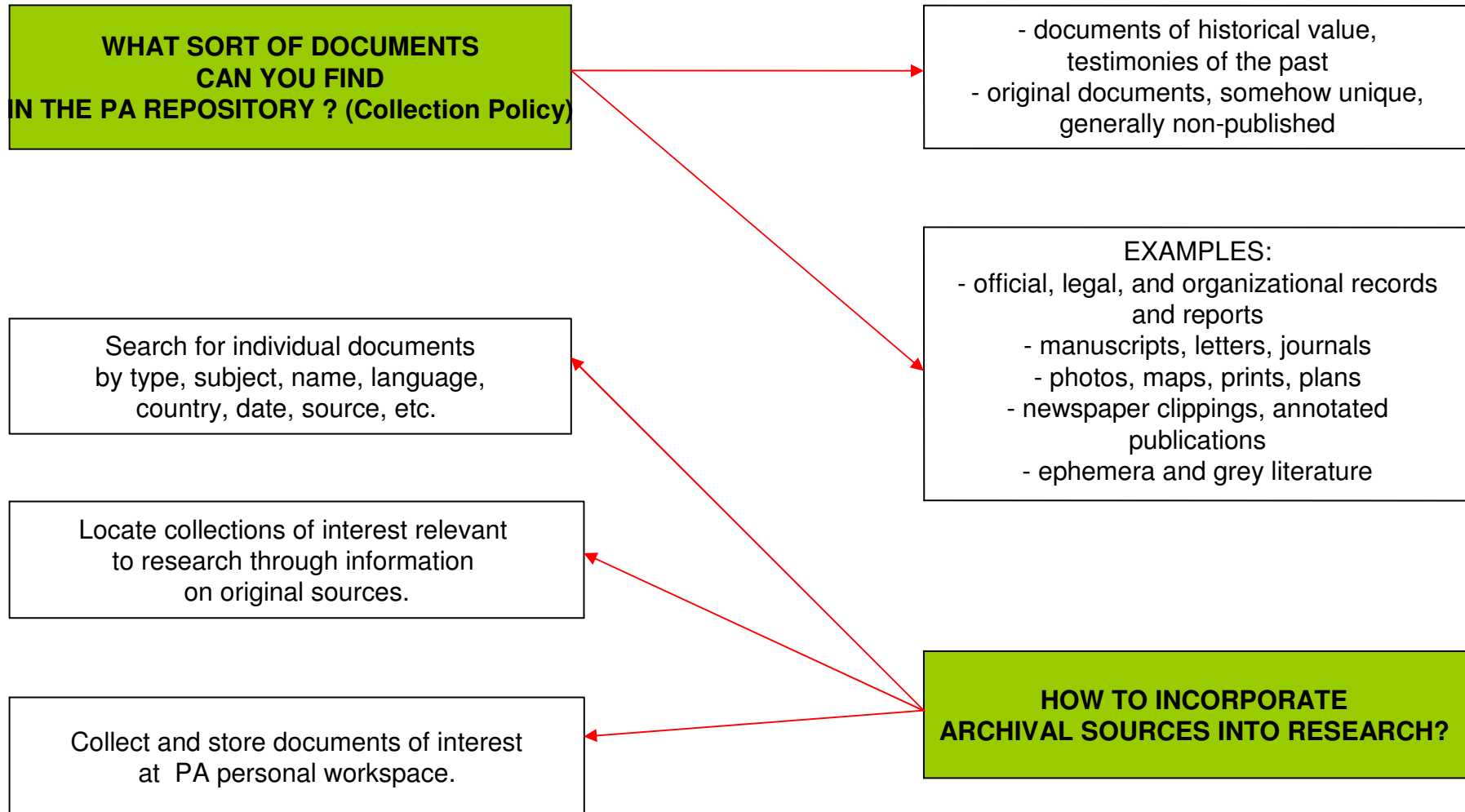


WHAT IS PARALLEL ARCHIVE ?

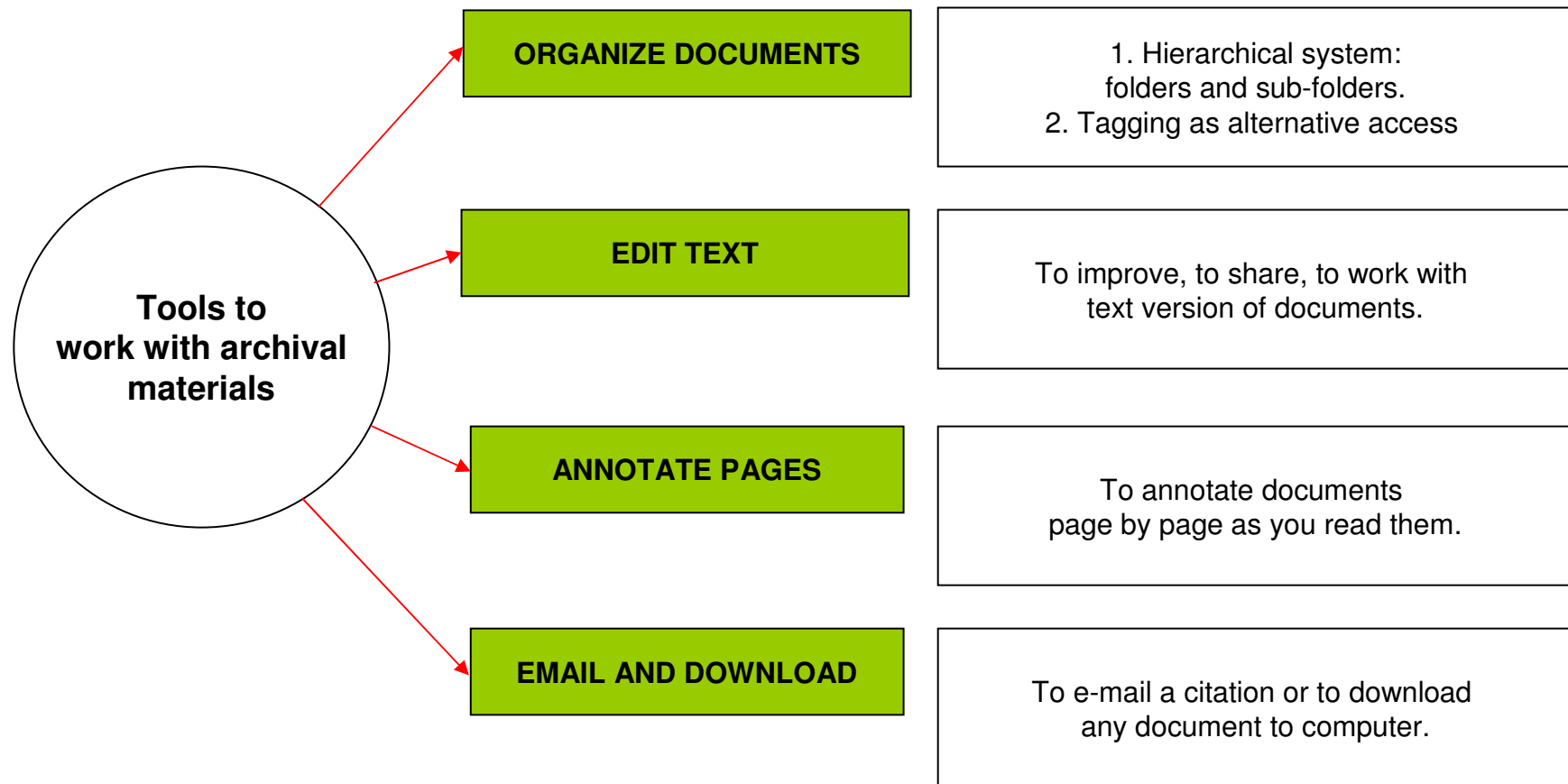
Parallel Archive is a space where researchers can upload, study, and share their scanned archival sources. It facilitates scholarly research, collaboration, and a critical approach to documents.



1 : A REPOSITORY TO FIND ARCHIVAL SOURCES



2 : USEFUL WORKING TOOLS FOR RESEARCH



CREATE A DOCUMENT ON PA (1)

Upload your document

PARALLEL ARCHIVE
beta

SEARCH »

Advanced Search

REPOSITORY

MyCollections MyShoebox MyTags MyForums MyProfile

UPLOAD

Standard Upload
Drag and Drop
Simple Upload

Upload

Click "Browse..." to select files. For multiple selection, please use the "Ctrl" or "Shift" key.

File	Size	Remove
------	------	--------

Accepted Formats: TIFF, JPG, PDF
Maximum File Size: 10MB
Maximum Files per Upload: 100
Temporary Storage Limit: 500MB

Need help?
Improve your Results

Browse... Upload

Browse and....

...upload your archival material

Choose
your
upload
method

CREATE A DOCUMENT ON PA (2)

Describe your document

Describe

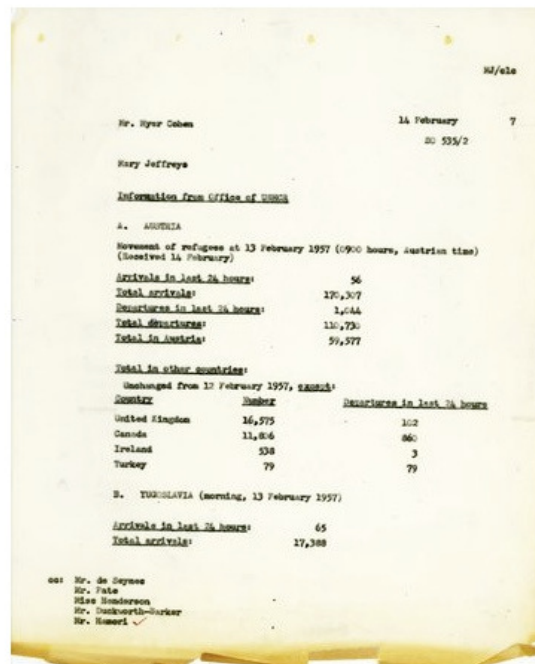
Enter information about your original document. English is strongly encouraged for all fields except Original Title.
Required fields are marked with a *.

« Prev 1 2 3 Next »

Zoom

+

Fullscreen



Original Document Type(s) * (select one or both)

☐ Image ☐ Text

Primary Language * (not required for image documents)

—

Additional Language(s) »

Original Title : Subtitle *

☐ given

English Title : Subtitle »

Original Document Creator(s) »

Other Contributor(s) »

Date of Original Document

Year * Month Day

☐ circa

If publication, provide Standard Citation »

Period Covered *

Year Year

Give as much information on the document as you can
and be sure to fill in all required fields !

The « i »
explains in
detail
how to fill in
each field

At the end of this process, you must :

- decide whether to keep your document private or make it public
- submit your document

3 : PERMANENT URLs FOR ACCESS AND CITATION

```
graph TD; A[3 : PERMANENT URLs FOR ACCESS AND CITATION] --> B[PERMANENT ACCESS TO DOCUMENTS]; A --> C[A TOOL FOR REFERENCE AND CITATION]; A --> D[CURATED SCHOLARLY CONTENT]; B --> E[Upload, store, study, and manage documents online anywhere through the internet]; C --> F[Provide direct links to primary sources in publications or papers]; D --> G[PA exercises curatorial control of content to maintain a scholarly standard];
```

PERMANENT ACCESS TO DOCUMENTS

Upload, store, study, and manage documents online anywhere through the internet

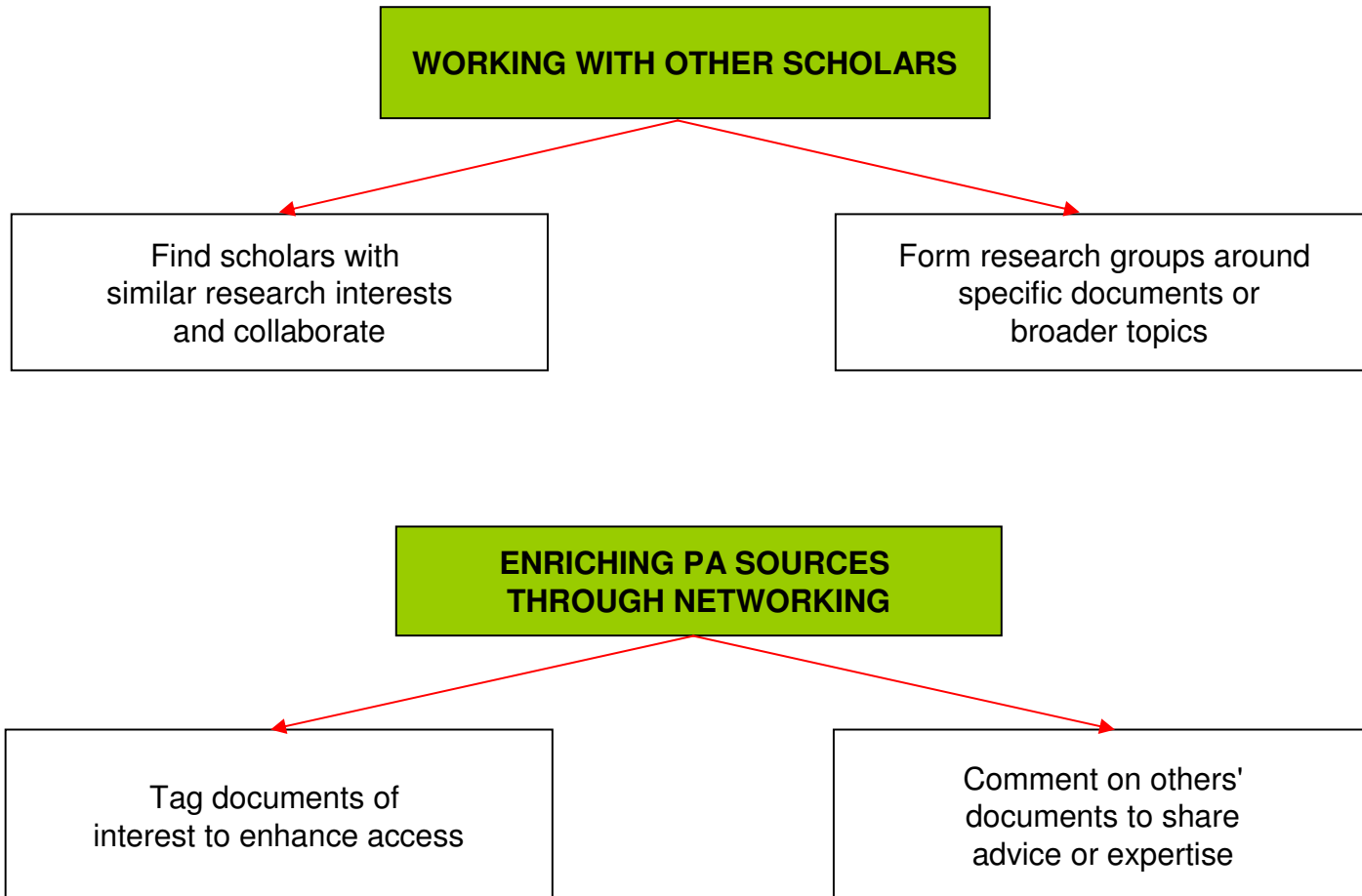
A TOOL FOR REFERENCE AND CITATION

Provide direct links to primary sources in publications or papers

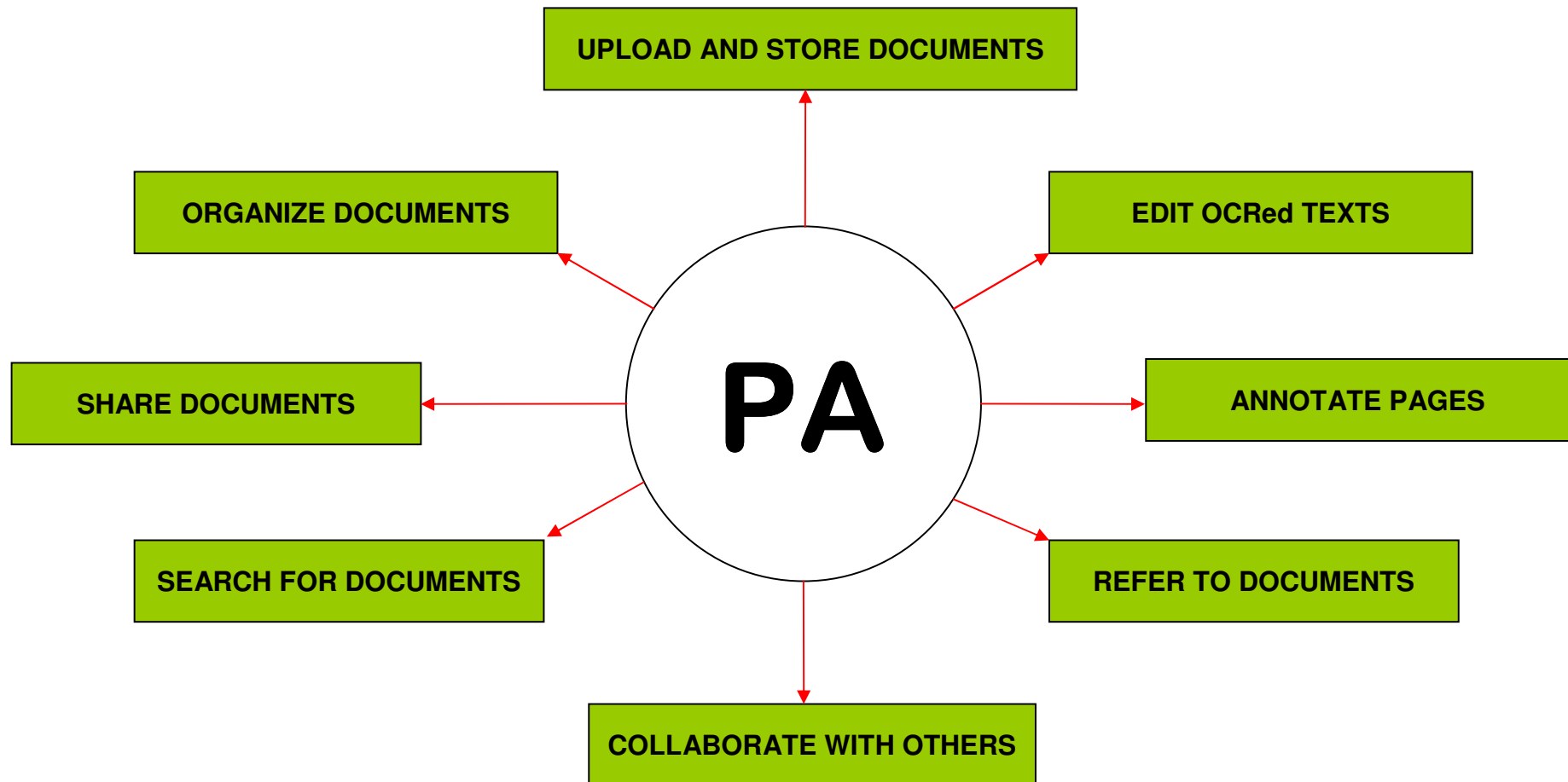
CURATED SCHOLARLY CONTENT

PA exercises curatorial control of content to maintain a scholarly standard

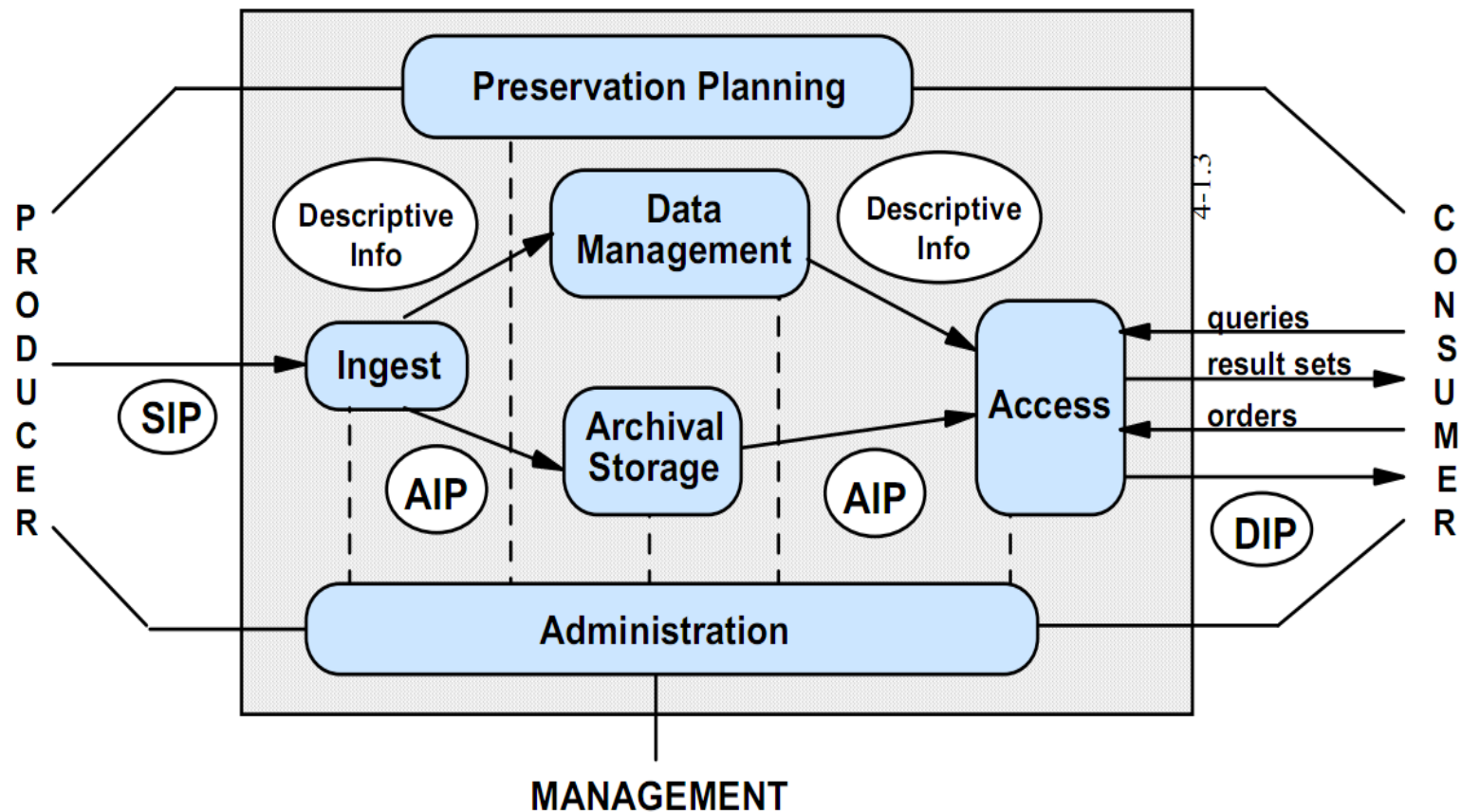
4 : A NETWORK OF SCHOLARS AND IDEAS



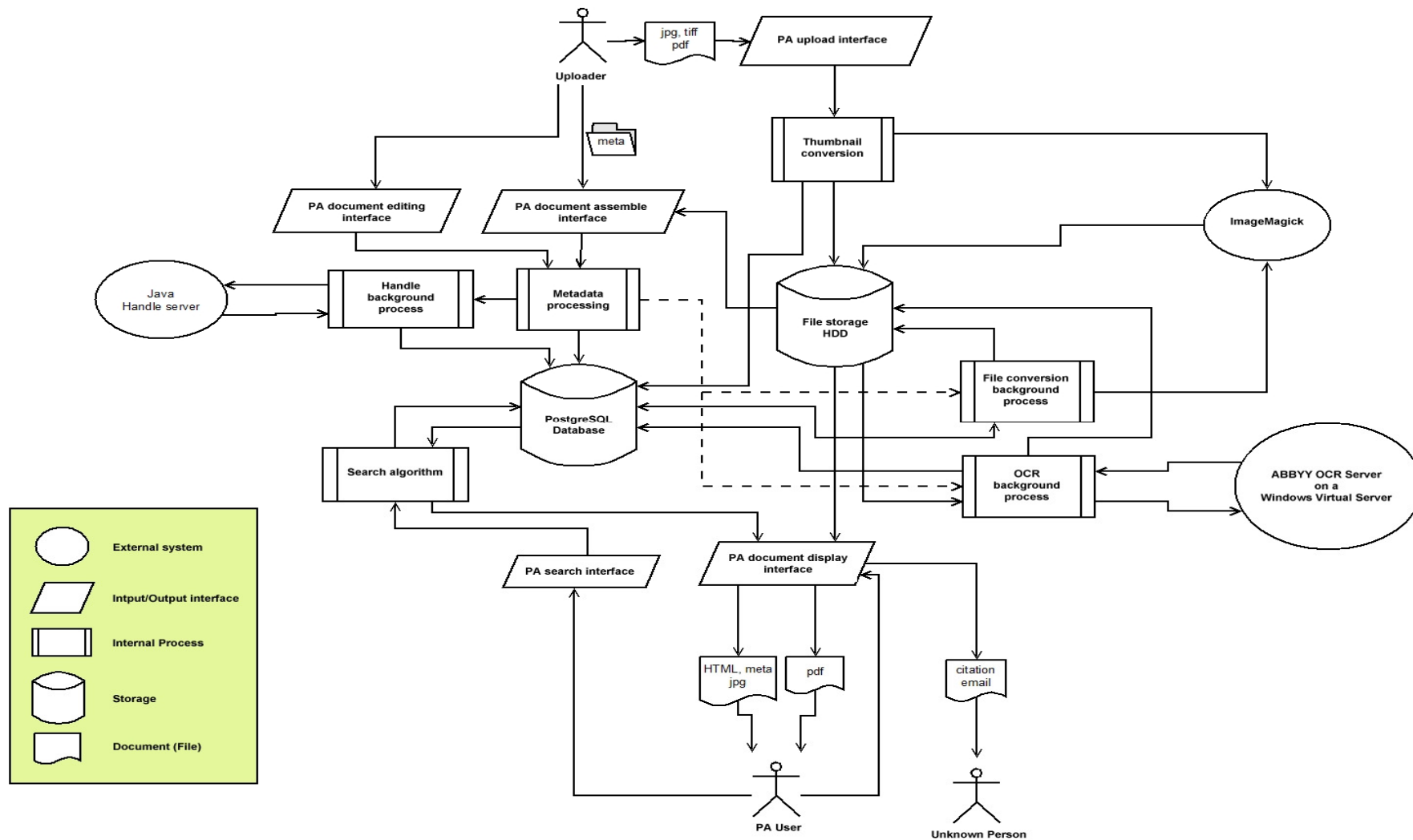
CONCLUSION : AT PRESENT A TOOL FOR INDIVIDUALS



1.1 Compliance with the OAIS Reference Model



1.2 PA Data Flow Diagram



2. Sustainability of digital formats and content in PA Context: preservation and curation

- **Digital formats, files:**
- *Original files: TIFF, JPEG, PDF*
- *Surrogates: JPEG, PDF, HTML*
- *Metadata: Dublin Core and Technical metadata saved*
- **Authenticity, provenance:**
- *Hash code is added to the original file*
- *Source is linked back to archives, contextual information is to be found*

- **Risks:**
- *Quality of master file is not good*
- *Preservation metadata schema needs to be developed*
- *Versioning is needed*
- *Master files need to be exported to other repositories*
- *The provenance is defined by the user, forgery?*
- *Archival finding aids are not item-level, digital object cannot be identified*

3. Sustainability of digital formats and content in PA Context: preservation and curation

- **Interoperability of hardware, software:**
 - *PHP programming language, disclosed source code*
 - *Open Formats: TIFF, JPEG, PDF*
 - *OS solutions except for OCR*
 - *PostgreSQL Database*
 - *Linux Red Hat configuration*
 - *Technical Documentation is almost ready*

- **Risks:**

- ***Eliminate dependency: OCR***
- ***Instead of Linux Red Hat, Debian GNU Linux?***
- ***Complete documentation!***

4. Sustainability of digital formats and content in PA Context: preservation and curation

- **Access:**

- *Unique persistent identifier, Handle.net System*
- *Private and public digital objects and metadata levels*

- **Digital Rights Management:**

- *Copyright Policy, Privacy Policy and Service, Code of Conduct*
- *Strong Moderation, procedures*
- *Collaboration with archives*
- *Planned: Legal Curation Center, Anonymisation Tool*

- **Risks:**

- *Privacy violation*
- *Copyright violation*
- *Violating restrictions by donors*

5. Sustainability of digital formats and content in PA Context: preservation and curation

- **Organizational Viability:**
- *OSA negotiates with a consortium of research institutes, archives to store master files in their repository*
- *Assigned PA team within OSA with clear roles*
- *Solid organizational background: OSA is part of the Central European University.*
- *Business Plan, Dissemination Plan, Fundraising Strategy is in progress*

- **Risks:**
- *Current data and content is not packaged yet to be preserved*
- *Small organization: lack of expertise, dependency on external developers*
- *Financial stability of the project*

6. Preservation and Curation Challenges in case of Parallel Archive

- **User driven system, not institution focused: "folksonomy"**
- **Dynamic content**
- **Dynamic system**
- **Authenticity**
- **The trusted repository model is hard to apply**
- **Data Management Plan Content Checklist: too theoretical**
- **"Mixed" digital content: quality of image, themes, metadata,**
- **Metadata enrichment: tagging, annotation, OCR, comments, forum posts**
- **Multilingual content**
- **What to preserve?**