

OPFCON 2020

Strengthening the French preservation community around formats and associated tools

Yannick GRANDCOLAS & Thomas LEDOUX
Bibliothèque nationale de France



Agenda

- *Building a community inside the BnF*
- *Going national with a « format watch unit »*
- *Sharing a registry of preservation tools*



{ BnF

- [illegible]

National watch unit on File formats

➤ Launch in 2019 of a national watch unit for file formats

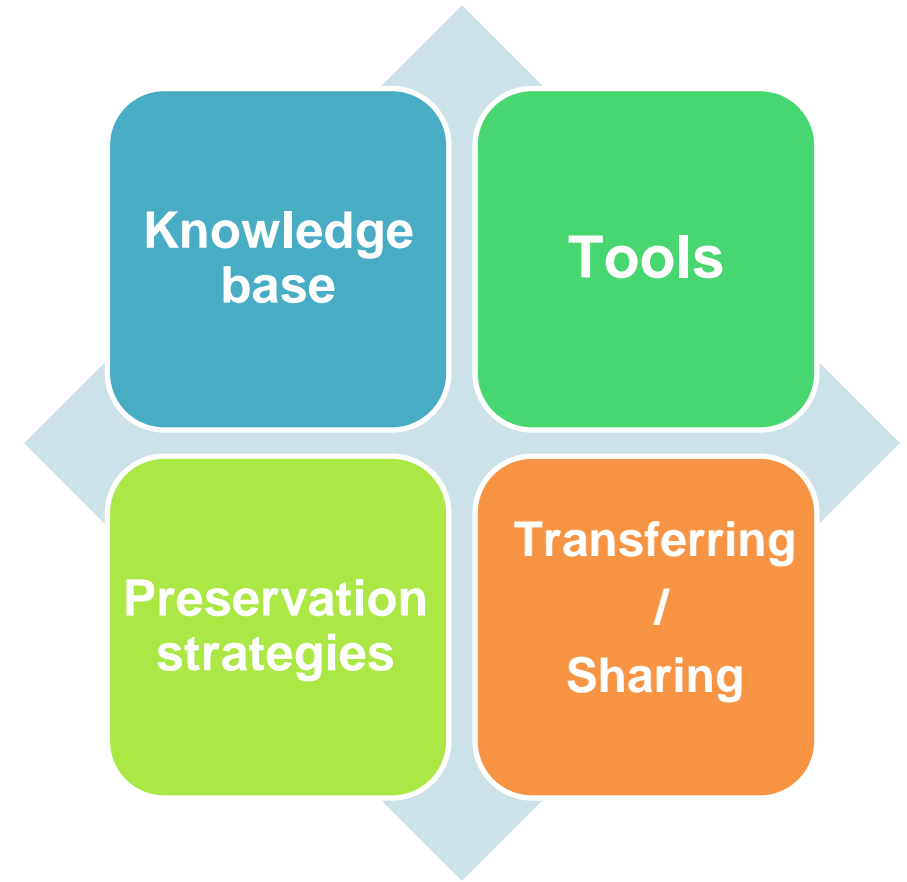
- Under the Aristote association and the working group PIN (*Pérennisation de l'information numérique* – Digital Information Preservation)
- About 10 partners (mainly, public national institutions)
- Open participation, need to be part of the Aristote association



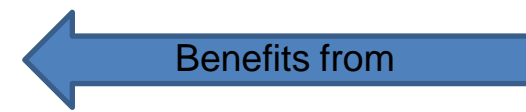
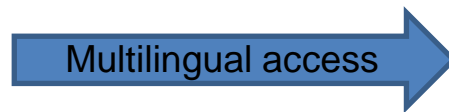
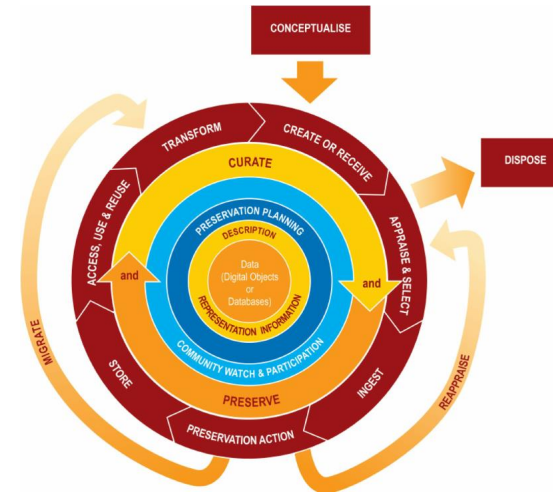
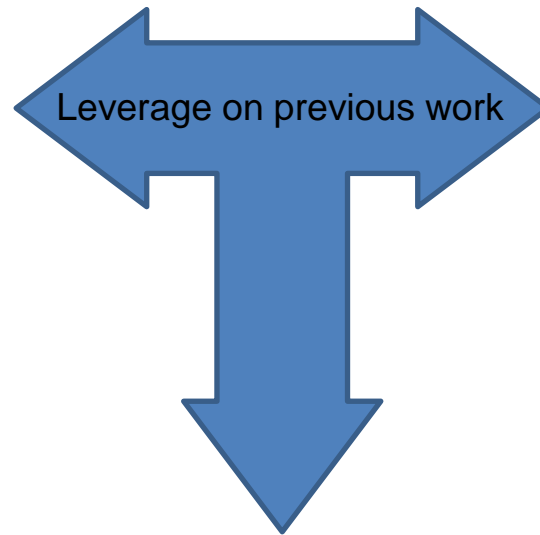
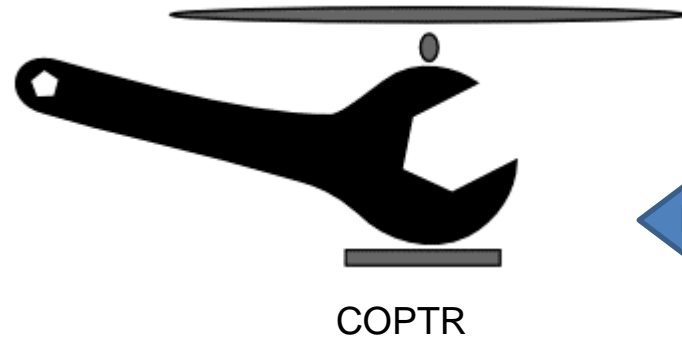
National watch unit on File formats

➤ Objectives

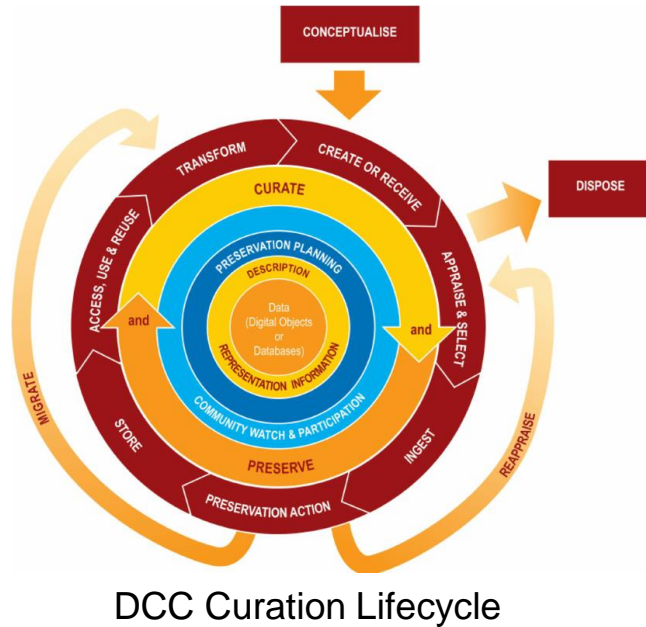
- Group watch activities
- Educate professionals on the subject
- Contribute and get involved in tools
- Be the French-speaking interlocutor at the international level



Tools registry



Step 1: create the categories of the registry



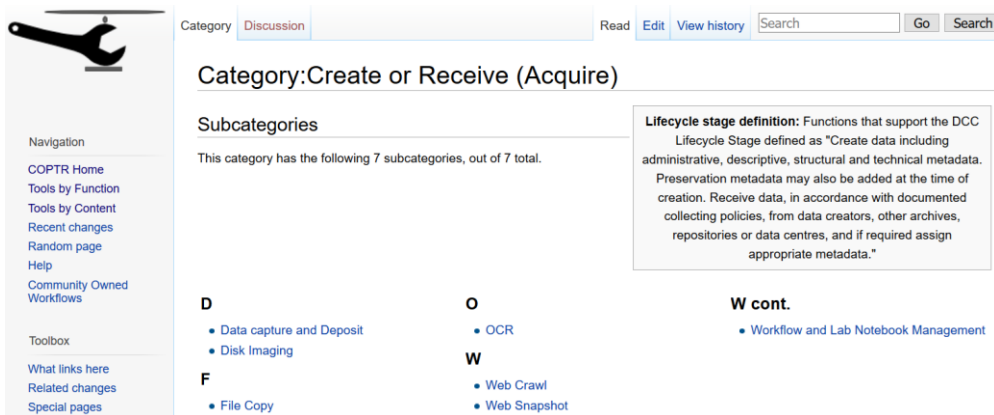
P31 (instance of - *nature de*) : Q714737 (category - *catégorie*)
 P361 (part of - *partie de*) : COPTR
 P1545 (series ordinal - *rang dans la série*) = 1

Each stage of the lifecycle becomes a WD entity

rang	qid	etape	stage
1	Q90878156	Accéder, Utiliser et Réutiliser	Access, Use and Reuse
2	Q90887043	Créer ou Réceptionner (Acquérir)	Create or Receive (Acquire)
3	Q89021580	Agir transversalement	Cross-Lifecycle Functions
4	Q90889758	Éliminer	Dispose
5	Q90885460	Verser	Ingest
6	Q90890255	Agir pour préserver	Preservation Action
7	Q90890401	Planifier la préservation	Preservation Planning
8	Q90890485	Stocker	Store

<https://w.wiki/N5r>

Step 2: links processes to the categories



Wikidata

P1269 (facet of - facette de) : ?stage

stage	pqid	process	processus
Create or Receive (Acquire)	Q4741970	file copying	Copie de fichier
Create or Receive (Acquire)	Q4929239	data collection	collecte de données
Create or Receive (Acquire)	Q61466324	web crawling	exploration du Web
Create or Receive (Acquire)	Q167555	optical character recognition	reconnaissance optique de caractères

Link the processes to the categories

Activities:

- Find the appropriate already existing entity
- Create a new one

Very often an entity for the result exists but not for the process itself for example : “disk image file” ([Q592312](#)) instead of “disk imaging”

<https://w.wiki/S4f>

Step 3: access to the tools



Wikidata

P31 (instance of - *nature de*)/P279(subclass - *sous-classe*)* Q7397 (software)

P366 (use - *usage*) : ?process

<https://w.wiki/S4q>

stage	etape	process	processus	tool	toolLabel
Ingest	Verser	file format identification	identification du format de fichier	Q wd:Q24305922	Digital Record Object Identification
Ingest	Verser	file format identification	identification du format de fichier	Q wd:Q16927990	JHOVE
Ingest	Verser	file format identification	identification du format de fichier	Q wd:Q2858088	Apache Tika
Ingest	Verser	file format identification	identification du format de fichier	Q wd:Q181820	file
Ingest	Verser	data integrity	intégrité	Q wd:Q6802825	Md5deep
Ingest	Verser	data integrity	intégrité	Q wd:Q1932013	Md5sum

Tools are classified
New tools are identified and discovered

Conclusions

- Work in progress:
 - Improve the classification
 - Find the appropriate processes or create new ones
 - Describe the tools in COPTR, not in Wikidata
 - Find a way to share with WikiDP

