

# OD ZPŘÍSTUPNĚNÍ ELEKTRONICKÝCH KVALIFIKAČNÍCH PRACÍ K JEJICH DLOUHODOBÉ ARCHIVACI

Mgr. Eliška Pavlásková, Ph.D.  
Ústav dějin a Archiv Univerzity Karlovy

10. ročník Konference o šedé literatuře a repozitářích, 19. 10. 2017

Prezentace je dostupná pod licencí Creative Commons, licence: [CC-BY-SA-4.0](https://creativecommons.org/licenses/by-sa/4.0/),  
via <http://www.nusl.cz/ntk/nusl-367304>



O elektroaffinitě aluminia.

Jaroslav Heyrovský.

č. 923.

## Uvod a rozvržení.

Přirodním cílem této práce byl potus o zjednodušení výroby čistého hydroxydu aluminia zavedením podobným Solwayově při výrobě louhu (návrh prof. F. J. Dornana). Že by se totiž aluminium z roztoku na rtuťové katodě vylučovalo, a takto vzniklá aluminiová amalgáma nechala se oxydovat, dávajíc přímo čistý hydroxyd.

Ža tím účelem bylo v první řadě nutno seznat, jak dalece je amalgámová elektroda aluminiová vůbec schopna existence a je-li to elektroda poratná.

Studie elektrolytického potenciálu této elektrody však předpokládá znalost koncentračních poměrů mezi ionty roztoků soli aluminiových. Proto jsem vyšetřil měřením elektrochemických sil

elektrody však předpokládá znalost koncentračních poměrů mezi ionty roztoků soli aluminiových. Proto jsem vyšetřil měřením elektrochemických sil

# Pozadí

- Kvalifikační práce na Univerzitě Karlově
  - 2006 – závěrečné kvalifikační práce vkládány do digitálního repozitáře
  - 2010 – odevzdávání prostřednictvím Informačního systému studium (SIS)
    - Workflow pro vkládání – tvorba SIP
- Zákon č. 499/2004 Sb., o archivnictví a spisové službě a o změně některých zákonů
  - *Dlouhodobé uložení digitálních archivních materiálů*
  - *Doporučení Národního archivu*

# Strategie dlouhodobé ochrany Formátová politika - zdroje

- Specifické potřeby Univerzity Karlovy
  - *Určená komunita*
  - *Tvůrci (původci) dat – studenti, zaměstnanci*
  - *Předběžná analýza odevzdaných prací*
- Doporučení Národního archivu
  - *Preferované formáty*
  - *Akceptované formáty*
  - *Formáty s nízkou trvanlivostí (např.. MS Word, interní formáty grafických aplikací...)*
- Dobrá praxe

# Dobrá praxe

## Ukazatele pro hodnocení formátů

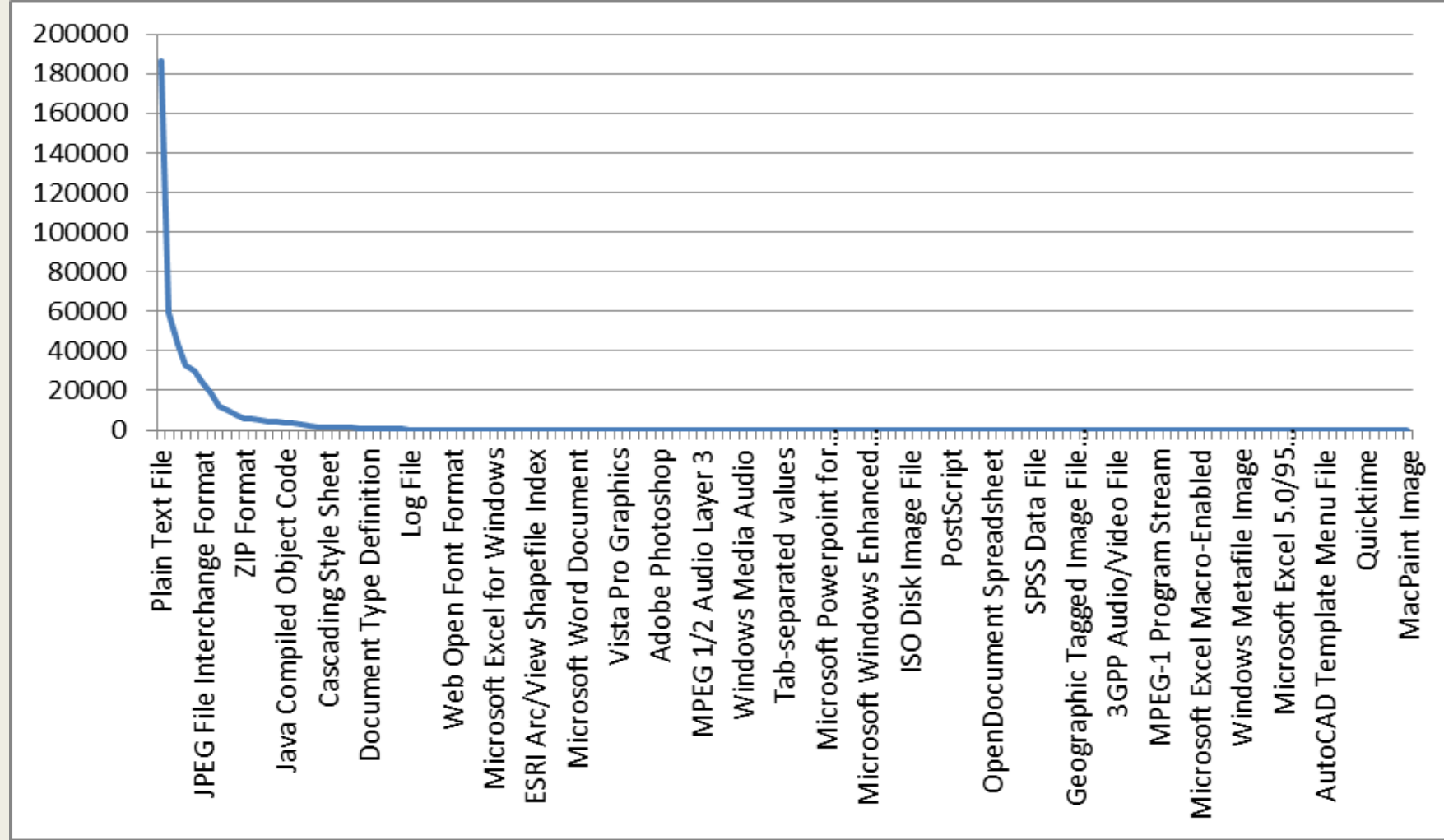
British Library	Library of Congress	MIT libraries	National Library of the Netherlands
Documentation and Guidance	Disclosure	Open, documented standard	Openness
Adoption and Usage	Adoption	Common usage by research community	Adoption
Complexity	Transparency	Standard representation (ASCII, Unicode) Unencrypted Uncompressed	Complexity
	Self-documentation		Self-documentation
External Dependencies	External dependencies		Dependencies
Legal Issues	Impact of patents	Non-proprietary	
Technical Protection Mechanisms	Technical protection mechanisms		Technical Protection Mechanism (DRM)
Development Status			
Software Support			
Embedded or Attached Content			
Other Preservation Risks			
			Robustness

# Předběžná analýza (přílohy)

- Leden 2015 - únor 2016
- 2 528 prací
- 481 396 souborů
  
- 148 rozdílných formátů (174 puid)
- Více než 20 obrazových formátů
- Více než 10 audiovizuálních formátů

Předběžná analýza

# Rozložení souborů vzhledem k formátům



## Předběžná analýza

# Nejvíce frekventované formáty

Format	Occurrence
Plain Text File	38,70%
Portable Network Graphics	12,28%
[not able to identify]	9,00%
Hypertext Markup Language	6,78%
JavaScript file	6,23%
JPEG File Interchange Format	4,98%
Extensible Markup Language	3,90%
Java language source code file	2,52%
Extensible Hypertext Markup Language	2,04%
Apple Double Resource Fork	1,70%
ZIP Format	1,26%
Acrobat PDF 1.5 - Portable Document Format	1,18%
Windows Portable Executable	1,04%



Formátová politika

# Struktura kvalifikační práce

Kvalifikační práce

Vkládá student

Text práce

Přílohy

Soubor

ZIP

Abstrakty (cze/eng)

atd...

Vkládá zaměstnanec

Posudek vedoucího

Posudek oponenta

Záznam o obhajobě

# Text práce

- born-digital
  - *Sázecí program LaTeX*
  - *Textové procesory*
- Nutná manuální kontrola konverze
- PDF/A – 1a, 2u
  - *PDF/A 3 – rizika z pohledu dlouhodobého uchovávání*

# Formátová politika

## Přílohy – schválené formáty

- Textové přílohy:
  - *PDF/A (verze 1a or 2u)*
- Obrazové přílohy:
  - *Joint Photographic Experts Group File Interchange Format (JPEG/JFIF, přípony: .jpeg, .jpg)*
  - *PDF/A (verze 1a or 2u)*
- Zvukové přílohy:
  - *Waveform audio format (WAV, \*.wav nebo \*.wave)*
  - *Moving Picture Experts Group Phase Audio Layer III (.mp3)*
- Audiovizuální přílohy:
  - *Moving Picture Experts Group Phase 2 (MPEG-2, \*.vob)*
  - *Moving Picture Experts Group Phase 4 (MPEG-4, \*.mp4)*
- Přílohy s charakterem dat (tabulky):
  - *Comma-separated values (CSV, \*.csv)*
  - *Extensible Markup language (XML, \*.xml) – balíček musí obsahovat relevantní XSD nebo DTD*
  - *Plain text file (\*.txt)*

## Formátová politika

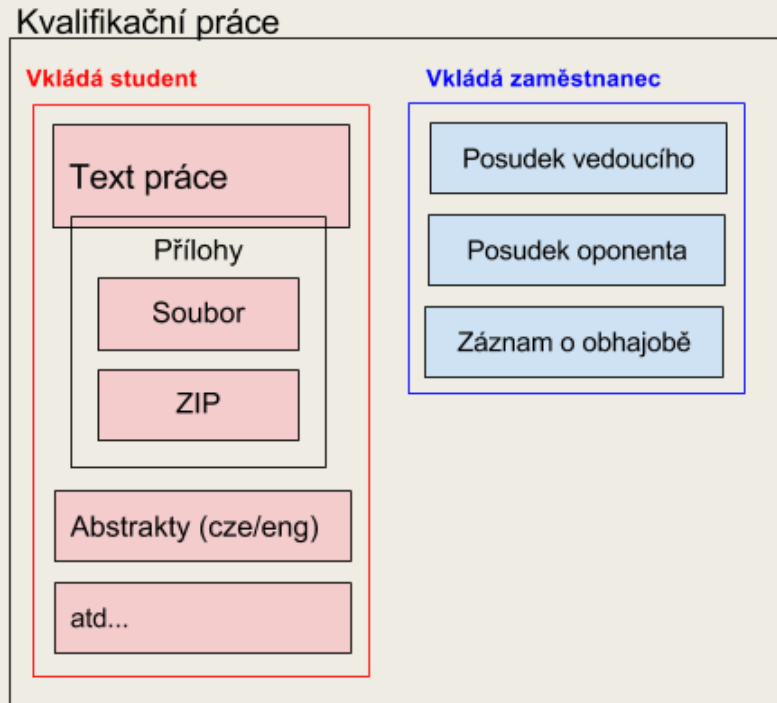
# Přílohy – neschválené formáty

### Doplňeny žádostí:

- a) název závěrečné práce
- b) jméno autora práce,
- c) důvod žádosti,
- d) formát přílohy (včetně verze nebo bližší specifikace) nebo programovací jazyk (včetně verze),
- e) aplikace potřebná k otevření nebo spuštění přílohy,
- f) operační systém (včetně verze), pod nímž lze přílohu otevřít nebo spustit,
- g) informace o licenci,
- h) seznam vkládaných souborů v nestandardním formátu,
- i) další relevantní informace (např. výčet a popis modulů a knihoven).

# Formátová politika

## Další dokumenty



### Abstrakty...

- PDF/A – 1a, 2u

### Posudky a záznam o obhajobě

- Born-digital nebo digitalizované
- PDF/A – 1a, 2u

# Pilotní fáze implementace

- Úpravy workflow pro odevzdávání
  - *VeraPDF – validační nástroj*
  - *Fido – identifikační nástroj*
  - *Formulář pro žádost o odevzdání přílohy v neschváleném formátu*
- Informační stránky
  - *Prozatímní návod na konverzi práce do PDF/A*
  - *Informace ke vkládání příloh*
- Elektronická poradna
- Březen 2017 – červen 2017
- Cca 3 000 prací a 5 000 souborů příloh

# Problematické oblasti

## Rozhraní pro odevzdávání

- VeraPDF
  - *Aktivní vývoj – komunikace s vývojáři*
  - *Nekonzistentní chování*
    - Verze
    - Parsery (PDFBox x Greenfield)
  - *Chyby při zpracování*
- Vlastní validační profil – sada testovaných vlastností
- Workflow odevzdávání
  - *Technické problémy*
  - *Problematika odevzdávání příloh*

# Problematické oblasti Konverze do PDF/A

- LaTeX
  - Mapování na unicode
    - *Obvykle chyba konverze*
  - Zpracování průhlednosti
    - *Obrázky*
    - *Grafy*
  - Porozumění návodu
- Verze PDF/A
    - *MS Word 2007 – PDF/A 1b*
    - *MS Word 2010, 2013 – PDF/A 1a*
    - *MS Word 2016 – PDF/A 3a*
    - *Libre Office – PDF/A 1a*
    - *Adobe Acrobat Professional*
  
    - *PDFCreator?*



# Vyhodnocení pilotního fáze

- Optimalizace workflow odevzdávání
  - *Asynchronní validace*
  - *Zobrazení validačního výstupu*
- Povolen formát Portable Network Graphics (png) pro obrazové přílohy
- Změna formuláře žádosti pro odevzdání přílohy v neschváleném formátu
- PDF/A nevyžadováno u dokumentů vkládaných zaměstnanci fakult

DĚKUJI ZA POZORNOST

[eliska.pavlaskova@ruk.cuni.cz](mailto:eliska.pavlaskova@ruk.cuni.cz)