

	Modul Upload Datasetov	Verzia :	1.2
		Dátum vydania :	08.12.2015
	Upload datasetov	Zodpovedný :	Tomáš Chovaňák

## 1.1 Upload datasetov

Tento modul zabezpečuje administráciu a manažment procesov ukladania rôznorodých datasetov do úložiska Google Storage prostredníctvom portálu Open Science.

### 1.1.1 Pridávanie datasetu

*Riešiteľ* : Tomáš Chovaňák

#### Analýza

Úlohou tohto modulu aplikácie je zabezpečiť proces uploadu datasetov od používateľa do globálneho úložiska. Dataset môžeme chápať ako jeden alebo viac súborov rôznych veľkostí a formátov spolu s metadátami. Keďže našou úlohou je zabezpečiť upload datasetu, znamená to upload viacerých súborov rôznych veľkostí a formátov spolu s ich metadátami.

Globálne úložisko datasetov je v našom prípade platforma ako služba Google Cloud Storage. Keďže datasetov môže byť v budúcnosti naozaj veľa, je potrebné ich ukladať do zrozumiteľnej hierarchie podľa dátumu vloženia.

#### Riešenie

	Modul Upload Datasetov	Verzia :	1.2
		Dátum vydania :	08.12.2015
	Upload datasetov	Zodpovedný :	Tomáš Chovaňák

### Popis sekvenčného diagramu :

1. Používateľ vyberá súbory, ktoré chce uploadovať. Na zabezpečenie správy procesu pridávania súborov a ich uploadu je použitá javascript knižnica <https://github.com/blueimp/jQuery-File-Upload>
2. Používateľ potvrdí spustenie uploadu.
3. Proces uploadu sa opakuje pre každý súbor.
  1. javascript klient posiela požiadavku metódou POST na server, ktorým iniciuje upload súboru.
  2. server sa autentifikuje aby mohol používať k Google Storage API. Autentifikačné údaje sú vytvorené vo vývojárskej konzole. <https://console.developers.google.com/project/openscience-paisou/apiui/credential>
  3. serveru je vrátený autorizačný token, server potom pošle POST požiadavku na Google Storage s názvom vytváraného objektu. Týmto sa inicializuje nový upload. V návratovej hodnote (objekt Response) sa nachádza upload id. Spolu s autorizačným tokenom je vrátený klientovi.
  4. Klient sa pomocou získaného tokenu a id uploadu vykoná PUT požiadavku v ktorej pošle súbor priamo na Google Storage.

Tento proces je popísaný na tejto stránke :

[https://cloud.google.com/storage/docs/json\\_api/v1/how-tos/upload](https://cloud.google.com/storage/docs/json_api/v1/how-tos/upload)

### Obmedzenie na veľkosť súborov :

Vzhľadom na to, že zároveň nechceme používateľa príliš zaťažovať zbytočnými akciami a tiež chceme aby boli upload a download datasetov čo najrýchlejší sme sa rozhodli pre nasledujúce obmedzenie :

- Ak sú pridané súbory datasetu dokopy väčšie ako 10 MB
  - Požadujeme od používateľa aby svoj dataset zabalil do jedného archívu alebo nejakým iným spôsobom zmenšil

Nechceme zaťažovať upload ešte aj automatickou kompresiou uploadovaných súborov. A tiež nechceme aby sa kompresia zbytočne vykonávala na strane servera niekedy po uploade.

### Hierarchia ukladania datasetov:

Datasety sú ukladané hierarchicky

`{rok}/{mesiac}/{nazov-datasetu}`

### Testovanie

Skúšobne bol vykonaný upload viacerých súborov s veľkosťou do 100 MB. Upload prebehol bez problémov a do modelu datasetu boli uložené v správnej forme aj všetky atribúty zadané používateľom (to sme overili v konzole <https://console.developers.google.com/project/openscience-paisou>).

	Modul Upload Datasetov	Verzia :	1.2
		Dátum vydania :	08.12.2015
	Upload datasetov	Zodpovedný :	Tomáš Chovaňák