



Evropská unie
Evropský sociální fond
Operační program Zaměstnanost



Návrh konceptu zavedení principů otevřených dat (Open Data)

Název projektu: Nastavení systému a regionální spolupráce v oblasti poradenství, poskytování dat a informací klientům TA ČR

Číslo projektu: CZ.03.4.74/0.0/0.0/15_025/0008941



Obsah

Úvod.....	3
1. Datové prostředí agentury.....	4
2. Mapování zveřejňovaných informací.....	5
3. Začlenění principů otevřených dat do prostředí agentury	6
4. Tvorba interních regulátorů	7
4.1. Publikace a katalogizace otevřených dat – procesní mapa	7
5. Praktické ověření	18
Přílohy	19
Příloha 1 – Podmínky užití otevřených dat.....	19
Příloha 2 – Text směrnice SME-38 Publikace a katalogizace otevřených dat v1.....	21



Evropská unie
Evropský sociální fond
Operační program Zaměstnanost



Úvod

Tento dokument si dává za cíl zmapovat prostředí Technologické agentury ČR (TA ČR) a definovat vhodná data pro zveřejnění v rámci Národního katalogu otevřených dat. Součástí dokumentu bude definování postupu publikování otevřených dat, určení rolí, které se budou procesu otevírání dat účastnit, tvorba interních předpisů a stanovení podmínek pro užití datových sad a pilotní ověření otevírání dat.

Dokument vznikl jako výstup projektu “Nastavení systému a regionální spolupráce v oblasti poradenství, poskytování dat a informací klientům TA ČR” CZ.03.4.74/0.0/0.0/15_025/0008941. Projekt je financován v rámci operačního programu Zaměstnanost z Evropského sociálního fondu a státního rozpočtu. Realizátorem projektu je Technologická agentura České republiky ve spolupráci s externím partnerem agenturou CzechInvest.



1. Datové prostředí agentury

Technologická agentura disponuje z titulu své činnosti zajímavými informacemi nejen pro odbornou, ale i laickou populaci. Rozhodování o směřování agentury, a tedy o směřování samotných dotací je v současnosti vždy podloženo interně zpracovávanými daty, jež pro tyto účely již z podstaty věci musí být kvalitní jak po obsahové, tak i po stránce technické a umožní jejich instantní jednoduché zpracování vedoucí k uceleným podkladům pro rozhodování vrcholných představitelů České republiky.

Valná většina dat týkajících se právě účelu podpory vědy a výzkumu je reportována do informačního systému VaVal, který nově disponuje standardním ovšem neveřejným rozhraním pro publikaci dat.

Technologická agentura vyvíjí vyhledávací aplikaci Starfos nad datovým skladem Dafos, kde jsou agregována data z více zdrojů. Primárním datovým zdrojem je v tuto chvíli právě IS VaVal, nicméně systém kombinuje navíc data patentová z evropského patentového úřadu a rozšiřuje komplexitu informací o subjektech v prostředí vědy a výzkumu.

Tato datová základna patří rozhodně mezi kandidáty na její uchopení v režimu otevřených dat v ideálním případě na úrovni linked open data.



2. Mapování zveřejňovaných informací

Při analýze zveřejňovaných informací na webových stránkách TA ČR bylo zjištěno, že dochází k periodickému generování PDF se statistickými údaji, které jsou pohledem právě do informací, jichž je TA ČR držitelem. Jejich strojové vytěžování není triviální a fakticky tak znemožňuje jejich přepoužití a zasazení do jiných kontextů právě například datovými vědci, či chceme-li, veřejností.

Při definování datových sad ke zveřejnění jsme se zabývali i otázkou, jež metodika Ministerstva vnitra k otevřeným datům také zmiňuje, a to zjednodušení přístupu k informacím podle 106/1999 Sb. Žádosti o informace v tomto režimu bývají v našem případě ovšem zpravidla přesně zacílená a standardně se odpovědi spíše budují z dostupných informací na míru každého dotazu. V tomto případě jsme tedy nějaké standardní strojově čitelné otevření nenavrhovali.

Zamýšlenými okruhy pro zveřejnění byly zejména tedy data:

- o podpoře aplikovaného výzkumu,
- dopady výzkumu,
- struktury účastníků výzkumu a
- hodnocení zacílení dotační podpory.

Nabízela se tedy možnost definice datových sad a úprav na straně informačních systémů agentury, ve kterých dochází k administraci podpory s cílem splnění podmínek míry otevřenosti dat na nejvyšších stupních.

Inspirací pro další otevřené datové sady byl (je a bude) web Technologické agentury, kde lze poměrně jednoduše dohledat další informace, které si o otevření v režimu otevřených datových sad vyloženě říkají.



3. Začlenění principů otevřených dat do prostředí agentury

Otevřená data jsou informace a data zveřejněná na internetu, která jsou úplná, snadno dostupná, strojově čitelná, používající standardy s volně dostupnou specifikací, zpřístupněná za jasně definovaných podmínek užití dat s minimem omezení a dostupná uživatelům při vynaložení minima možných nákladů. Konkrétně jde o různé statistiky, rozpočty, přehledy, databáze apod.

Od začátku jsme nepřístupovali k samotnému uchopení aktivit spojených s otevíráním dat jako k jednorázové činnosti, která pouze čistě splní legislativní povinnosti, ale vnímáme zabezpečení problematiky jako soubor kontinuálních činností a postupů, kdy se při zveřejňování jakýchkoliv informací bude posuzovat, zda-li by nebylo moudré informace zpřístupnit v režimu otevřených dat, či přímo nezměnit přístup k datům obecně tak, že by se nezveřejňovali statické informace ve strojově nečitelném formátu, ale vybudováním datových rozhraní jednotlivých sad by pak bylo možné interaktivně s využitím dostupných nástrojů data vizualizovat, ale zároveň i poskytovat v otevřené formě ideálně v režimu linked open data.

Toto jistě není možné zabezpečit bez adekvátní podpory managementu organizace a bez podpory datových analytiků, specialistů IT a dalších profesí fungujících v rámci agentury a je nutné ustavit do interních regulátorů odpovídající předpis a do popsaného prostředí zasadit i procesy týkající se otevřených dat. Zabezpečením periodických přezkumů a schvalování publikačního plánu by také mělo dojít k tomu, že problematika otevřených dat tzv. nezapadne.



4. Tvorba interních regulátorů

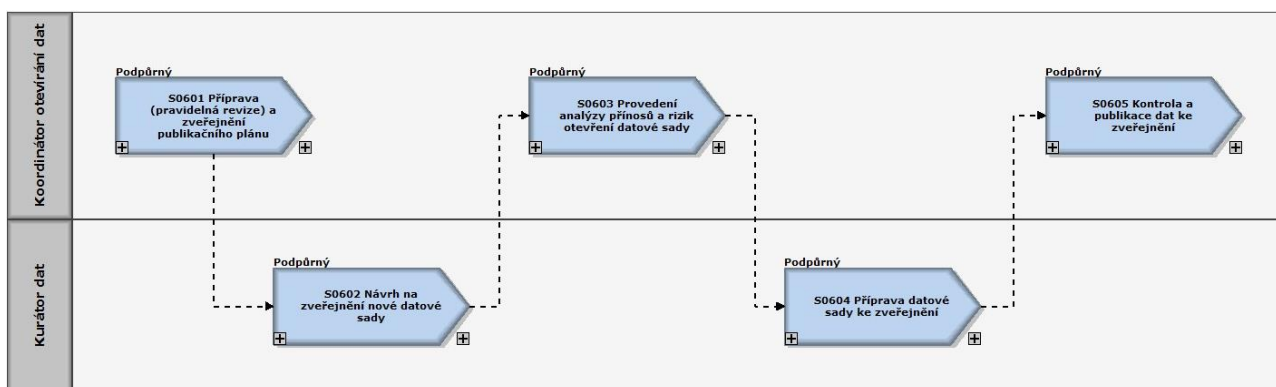
Při tvorbě interní směrnice jsme vycházeli z metodiky Ministerstva vnitra a vzorů poskytovaných na opendata.gov.cz. Výsledná směrnice a proces jsou uzpůsobené tak, aby bez problémů zapadla do fungujícího normativního, procesního a technického rámce agentury a využívají dostupné interní zdroje, a to jak technické, tak i organizační (analytický úsek, role bezpečnostního ředitele, právní oddělení, a další). Ve spolupráci s právním oddělením také vznikly základní podmínky používání otevřených dat TA ČR – dostupné v lokálním katalogu otevřených dat (jsou přílohou tohoto dokumentu).

Jmenováním koordinátora otevřených dat zabezpečíme pravidelný styk světa otevřených dat s prostředím managementu agentury a zároveň dojde ke koncentraci přehledu o datovém prostředí agentury u jedné role, což by mělo odbourat možný budoucí nekoncepční přístup k otevřeným datům a ke zpracování dat obecně.

4.1. Publikace a katalogizace otevřených dat – procesní mapa

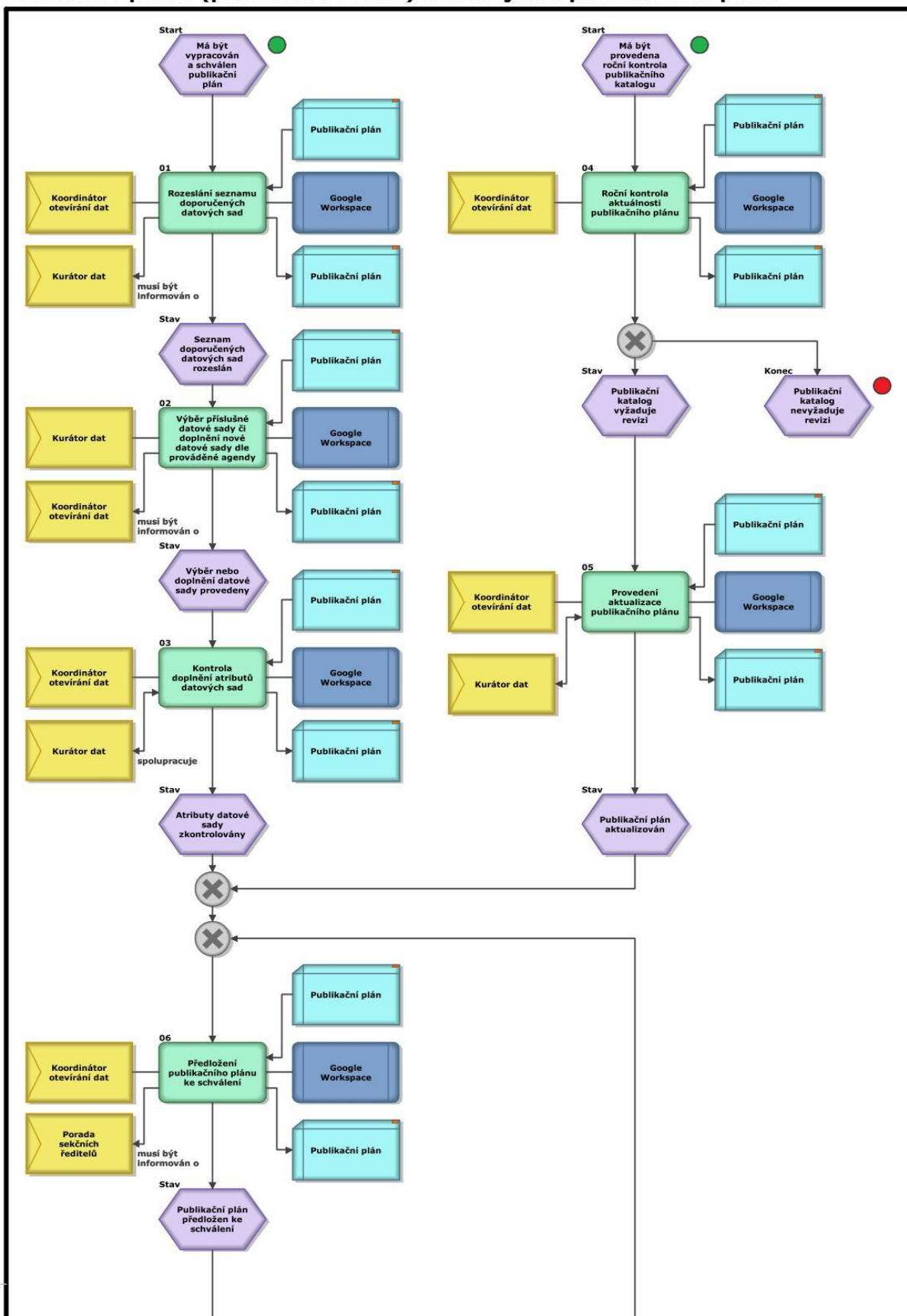
Procesní mapa

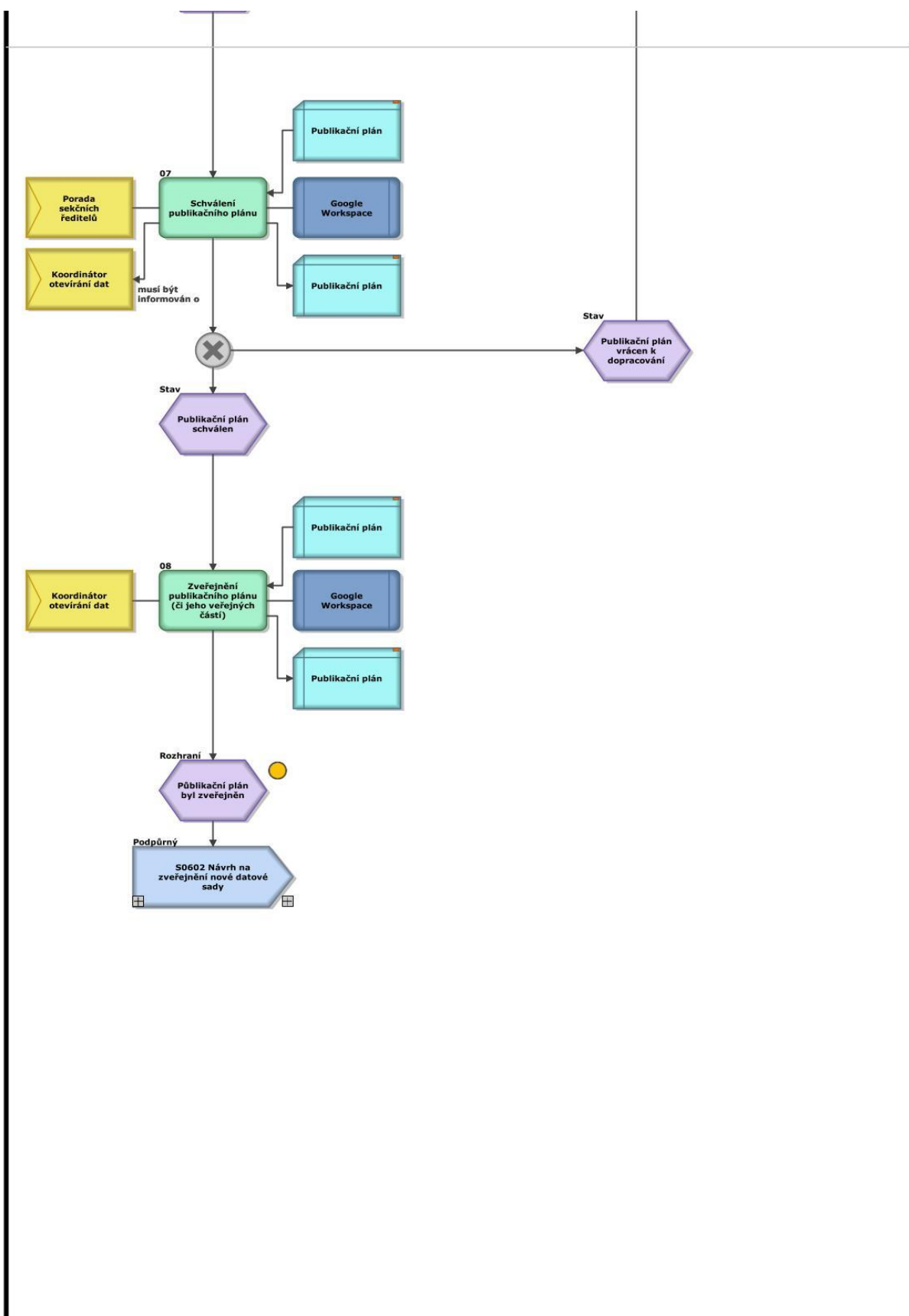
- S0601 Příprava (pravidelná revize) a zveřejnění publikačního plánu
- S0602 Návrh na zveřejnění nové datové sady
- S0603 Provedení analýzy přínosů a rizik otevření datové sady
- S0604 Příprava datové sady ke zveřejnění
- S0605 Kontrola a publikace dat ke zveřejnění





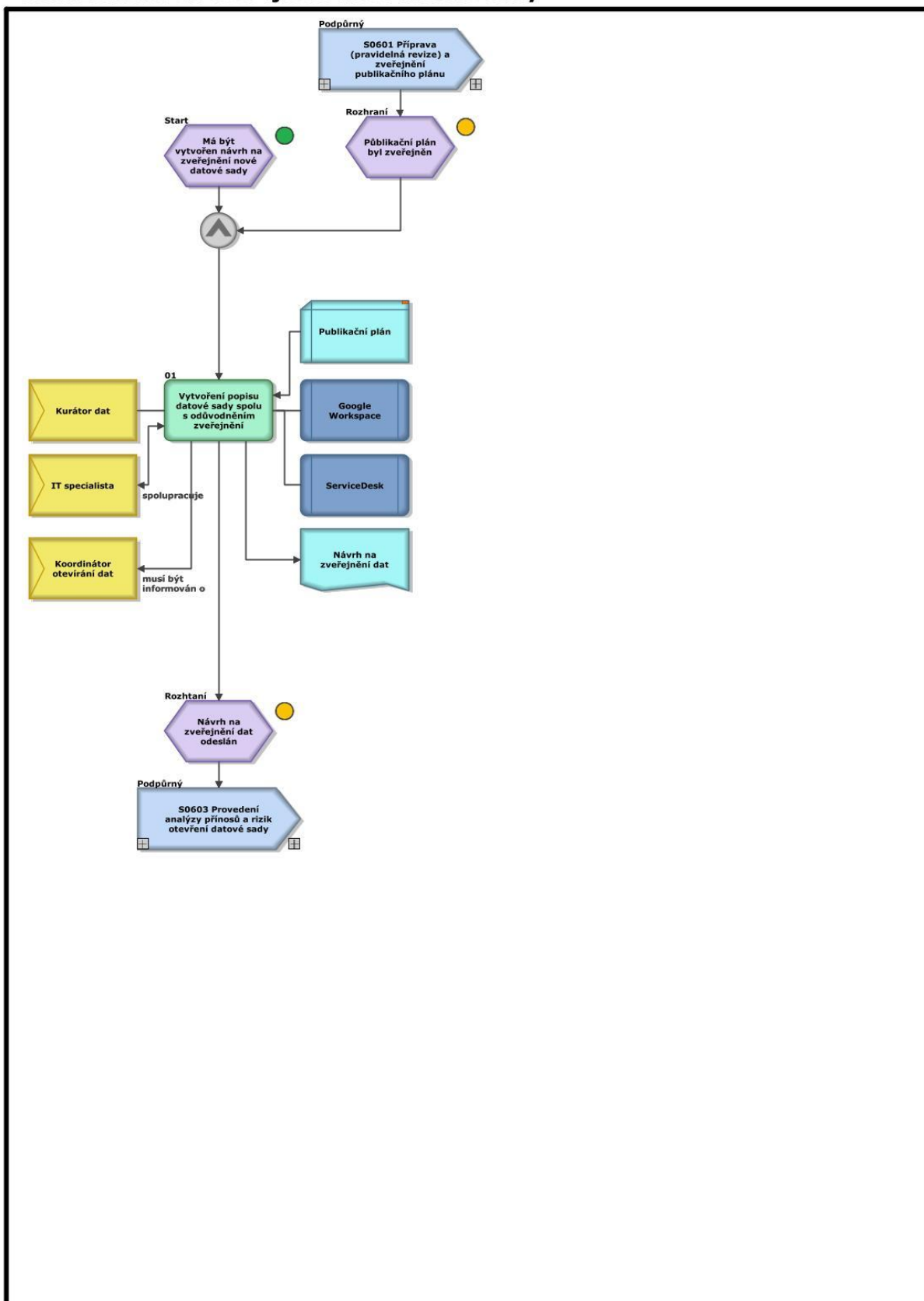
S0601 Příprava (pravidelná revize) a zveřejnění publikačního plánu





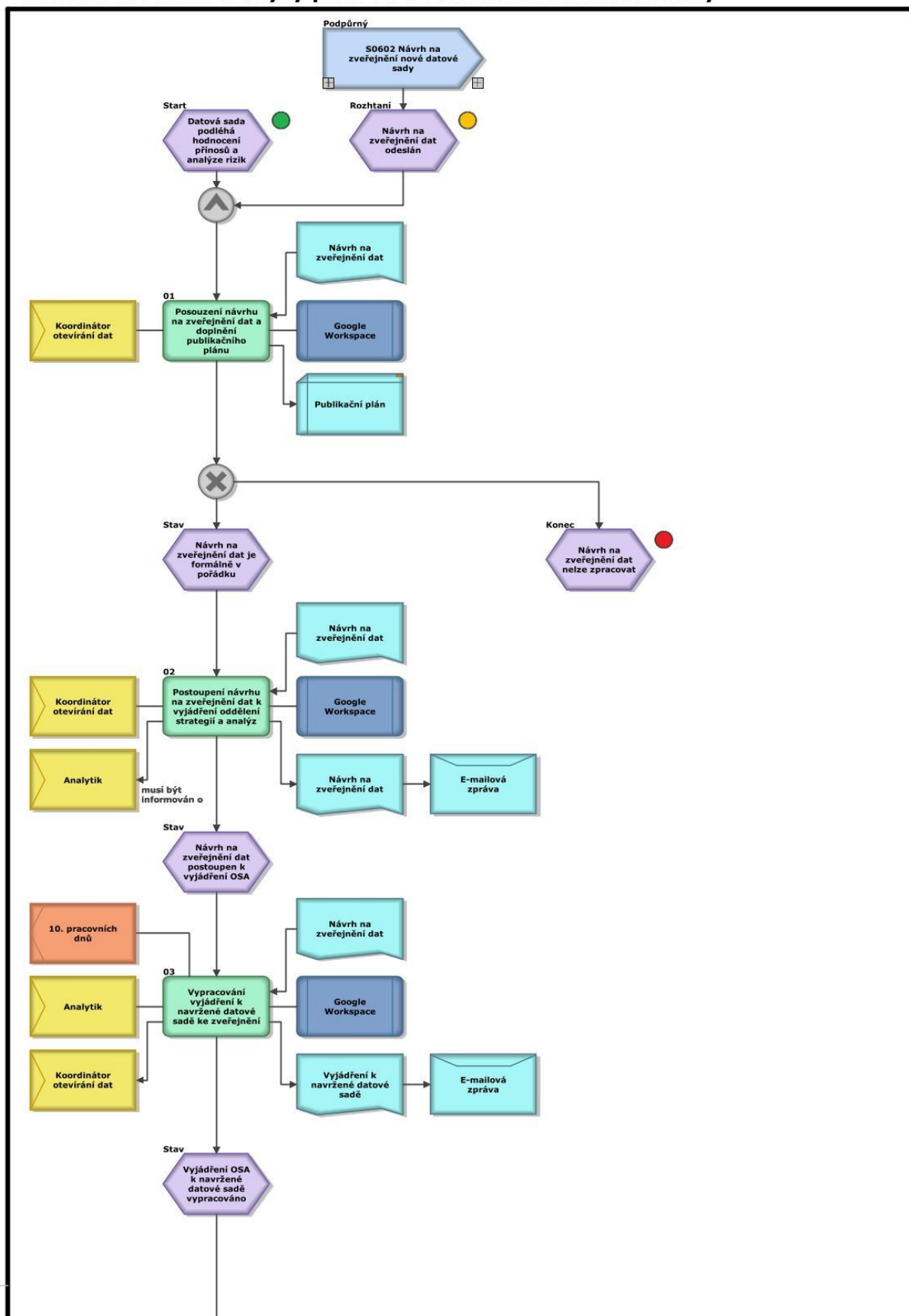


S0602 Návrh na zveřejnění nové datové sady





S0603 Provedení analýzy přínosů a rizik otevření datové sady

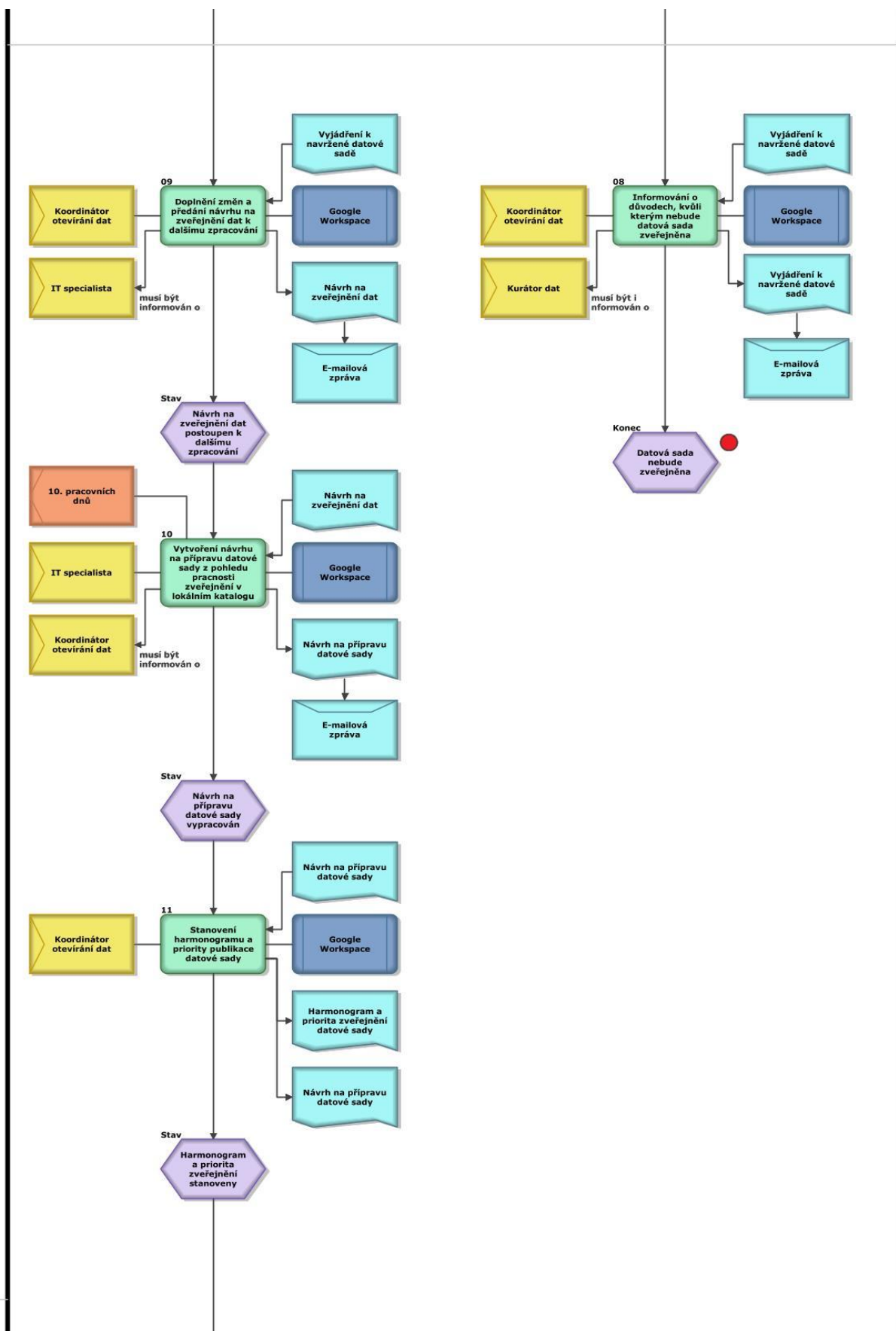


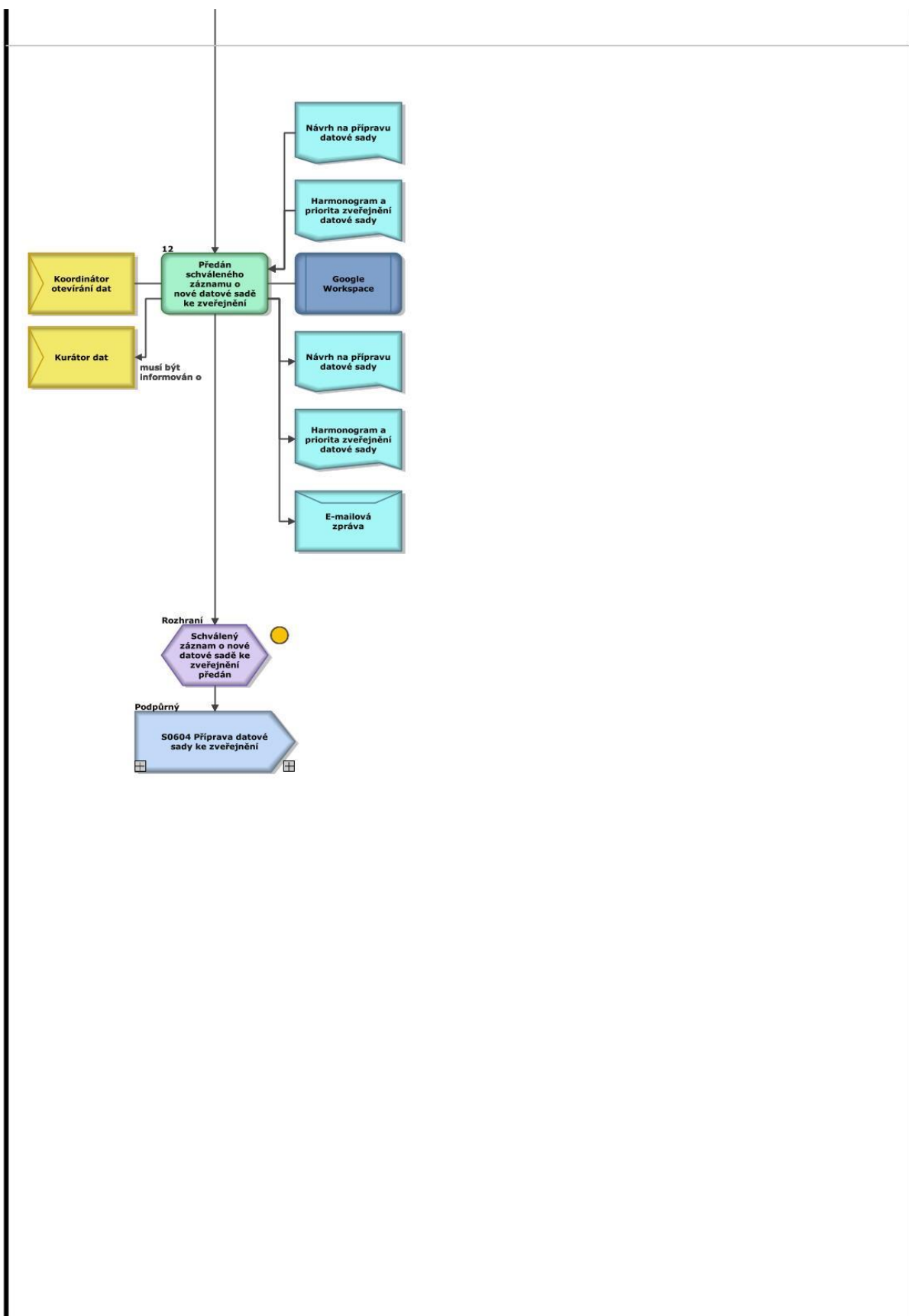
T A
Č R



Evropská unie
Evropský sociální fond
Operační program Zaměstnanost

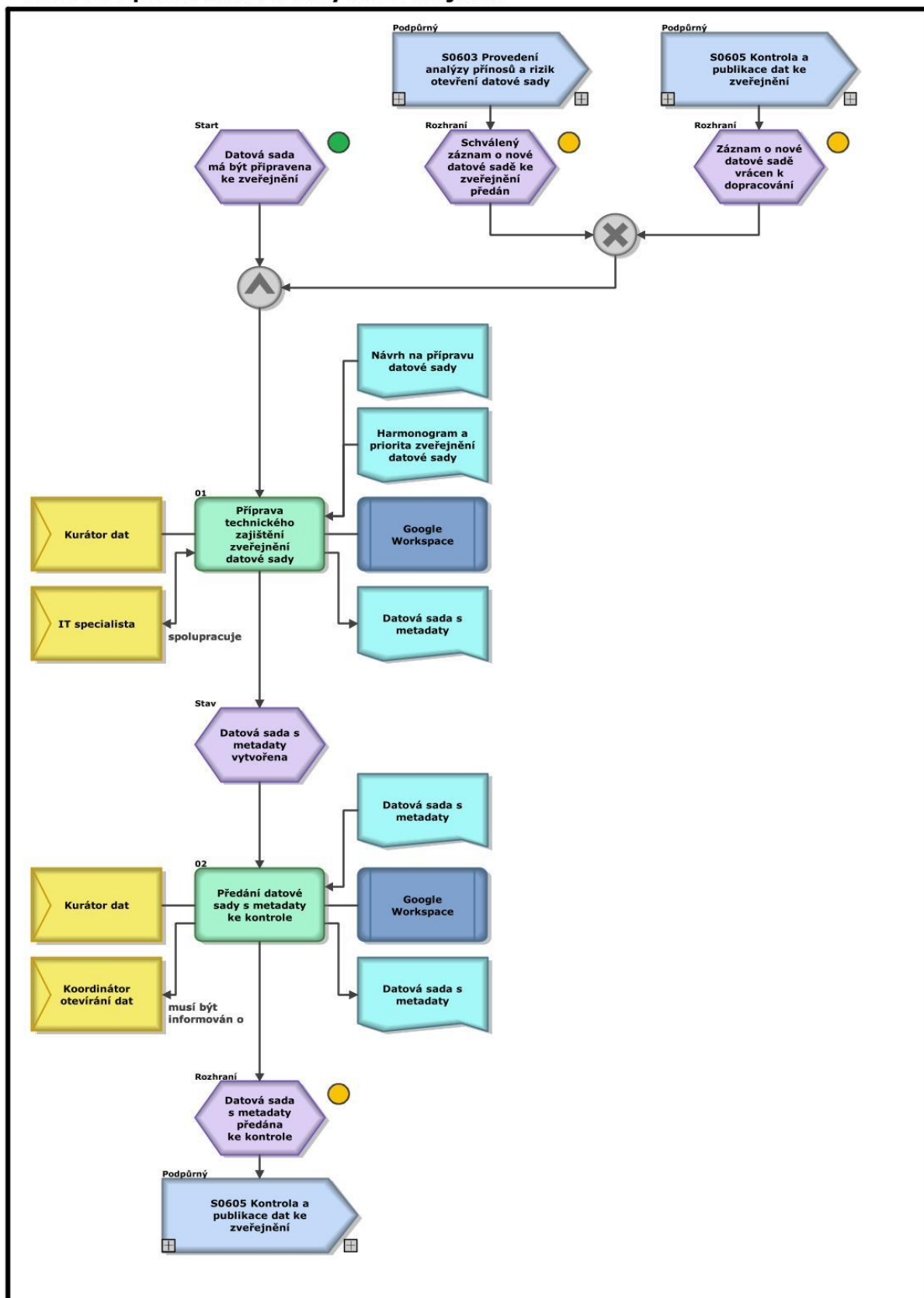
T A
Č R





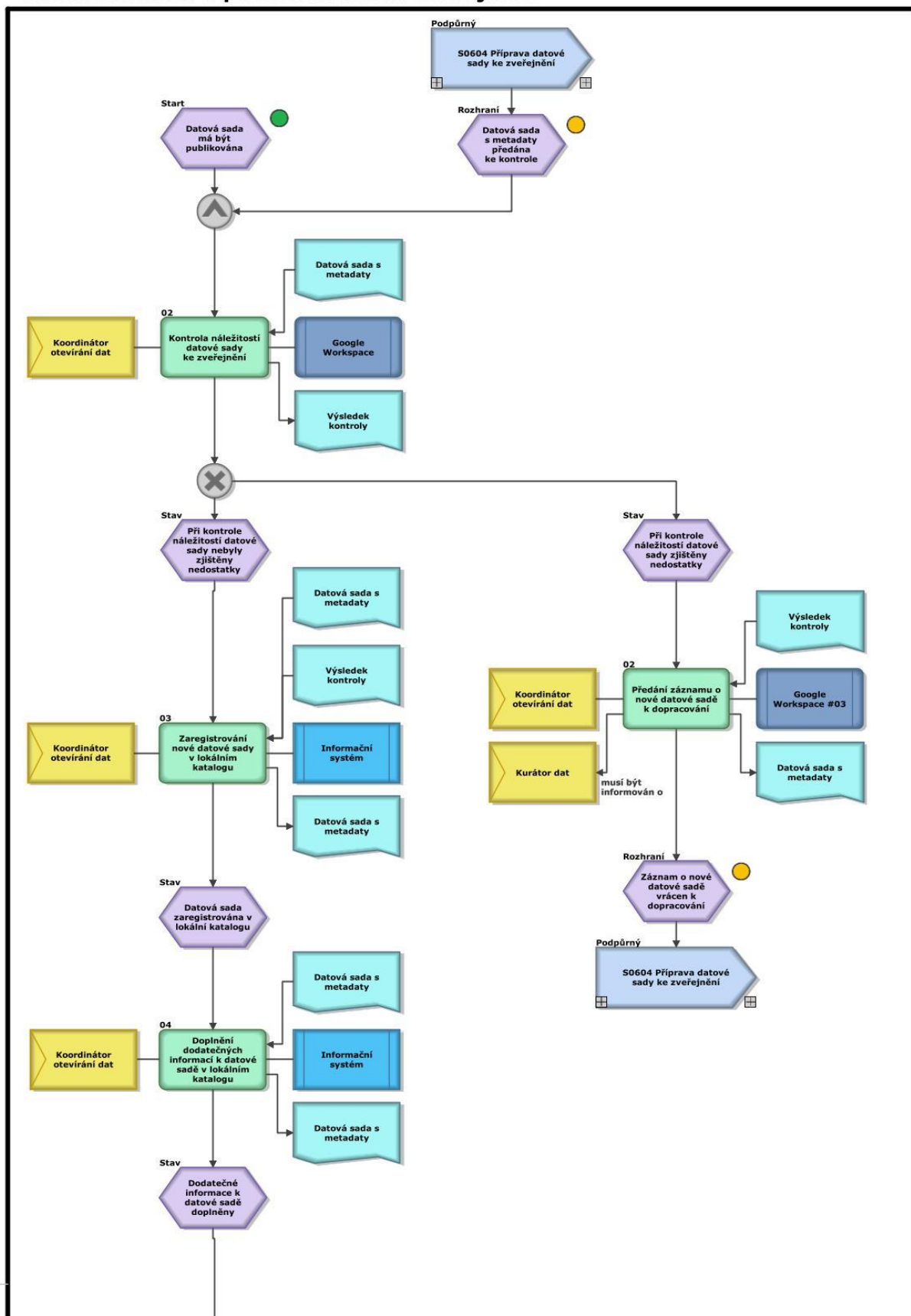


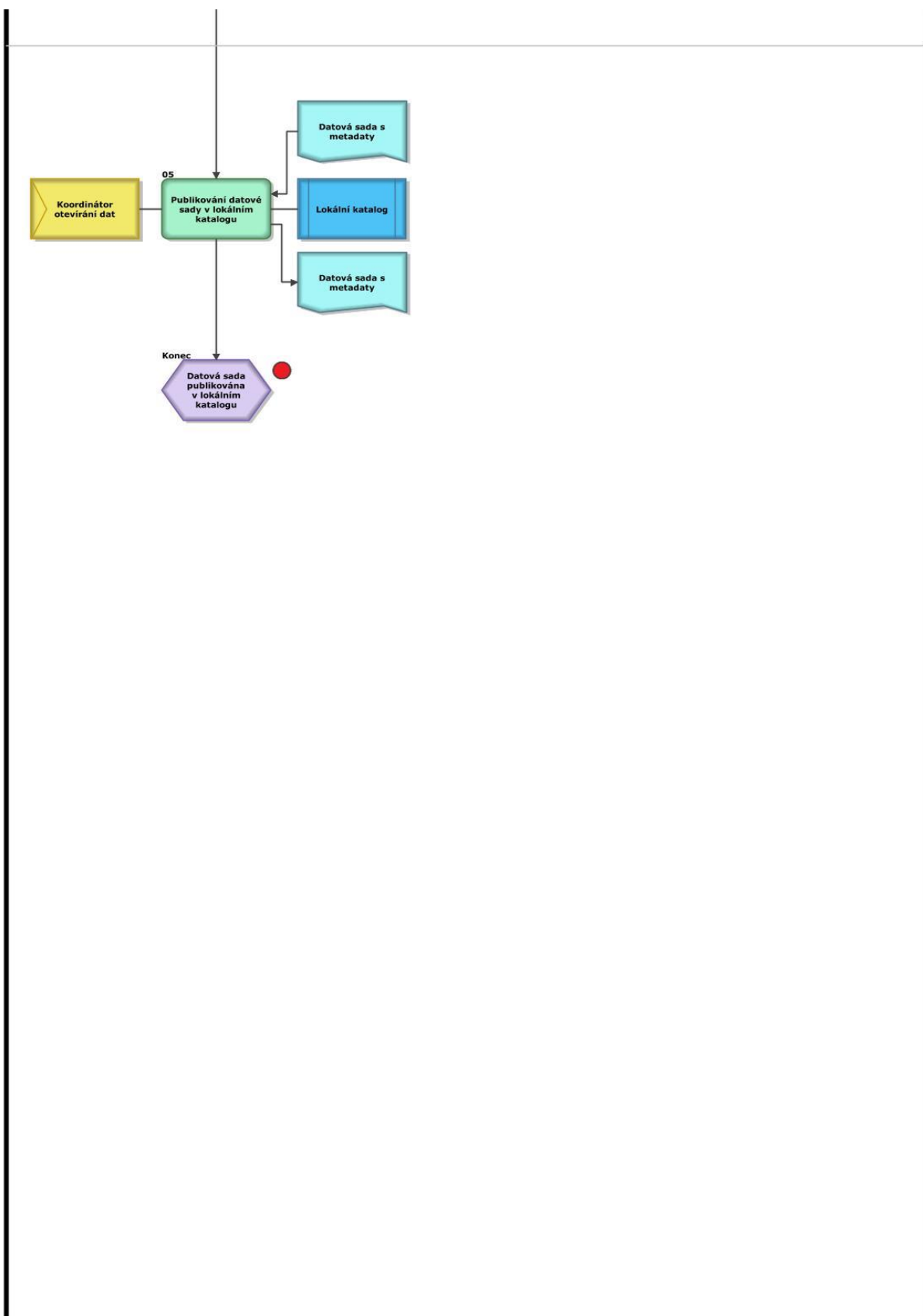
S0604 Příprava datové sady ke zveřejnění





S0605 Kontrola a publikace dat ke zveřejnění

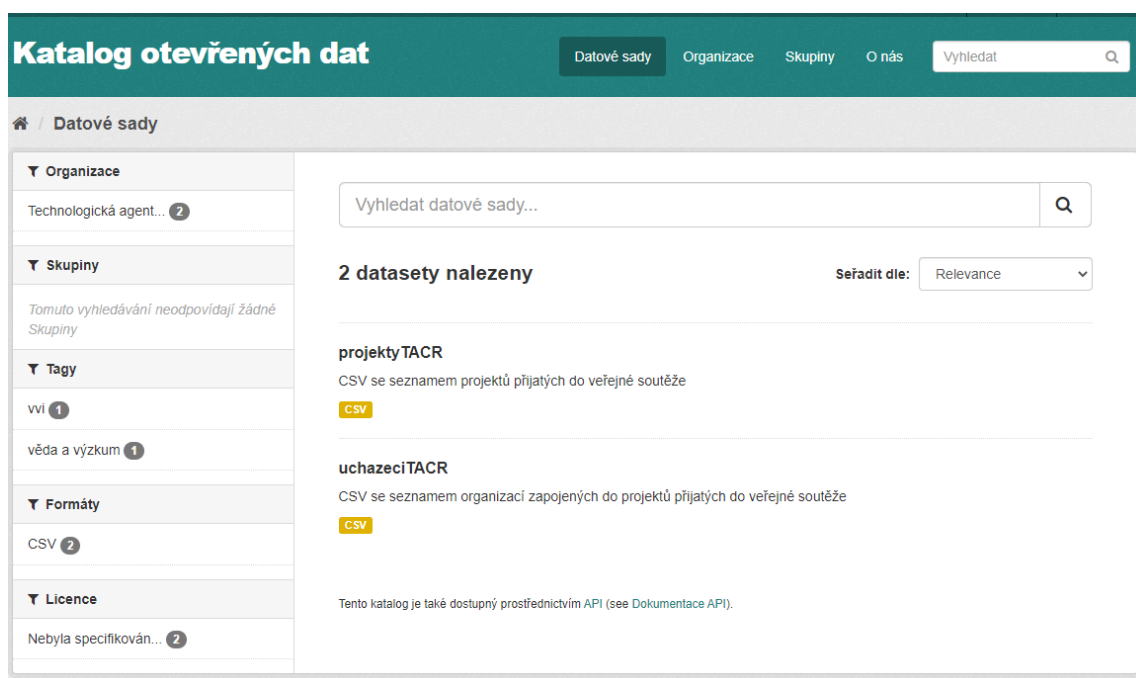






5. Praktické ověření

S tvorbou směrnice (SME-38 Publikace a katalogizace otevřených dat v1) docházelo zároveň k pilotnímu technickému a organizačnímu ověření definic uvedených v předpise samotném. Na straně informačního systému ISTA byly vytvořeny pilotní datové sady vyhovující principům otevřených dat. Ověřili jsme si postup registrace datové sady do lokálního katalogu a v testovacím provozu byla nasazena instance lokálního katalogu TA ČR.



Dochází k mapování zveřejňovaných informací a ke konstrukci první iterace publikačního plánu. Zároveň dochází k zadokumentování nově poskytované ICT služby, jež v důsledku povede k instalaci lokálního katalogu otevřených dat do technologické infrastruktury agentury. Registraci zmapovaných/vytvořených datových sad do lokálního katalogu otevřených dat bude následovat registrace lokálního katalogu do národního katalogu otevřených dat, čímž se uzavře kompletní první fáze samotné implementace řešení problematiky otevřených dat na TA ČR. Kurátoři dat budou instruováni, aby vždy při zveřejňování informací neopomíjeli možnosti otevření datových sad a řešili pak samotné provedení již ověřeným popsáním procesem otevírání datových sad.

V současné době nepředpokládáme technologické propojení lokálního a národního katalogu – registrace datových sad v národním katalogu tak bude probíhat prozatím manuálně.



Přílohy

Příloha 1 – Podmínky užití otevřených dat

Data Technologické agentury ČR (dále jen „TA ČR“), která jsou zveřejněna na webových stránkách TA ČR (<http://www.tacr.cz>) / v části Katalog otevřených dat TA ČR a která jsou označena jako otevřená data (dále jen „otevřená data TA ČR“):

- nejsou způsobem výběru nebo uspořádáním obsahu vlastním duševním výtvorem autora ve smyslu § 2 odst. 2 zákona č. 121/2000 Sb., o právu autorském, o právech souvisejících s právem autorským a o změně některých zákonů, ve znění pozdějších předpisů (dále jen „Autorský zákon“), případně jsou úředním dílem ve smyslu § 3 písm. a) Autorského zákona;
- nenaplnují znaky pro vznik práva pořizovatele databáze dle § 88 a následujících Autorského zákona;
- v sobě neobsahují autorským právem chráněná díla dle § 2 odst. 1 Autorského zákona.

Z těchto důvodů nejsou tato otevřená data z hlediska autorského práva a sui generis databázových práv chráněna jako celek, stejně jako nejsou chráněny jejich dílčí části. Otevřená data TA ČR tak smí uživatel:

- šířit – kopírovat, distribuovat a sdělovat veřejnosti,
- využívat a citovat v jiných dílech,
- využívat komerčně,

a to při dodržení následujících podmínek:

- uživatel je povinen vždy uvést dostatečně zřetelným a srozumitelným způsobem, že zdrojem dat je TA ČR, nikoli však takovým způsobem, aby vznikl dojem, že TA ČR podporuje uživatele dat nebo způsob, jakým jsou otevřená data TA ČR využita,
- uživatel není oprávněn otevřená data TA ČR jakkoli upravovat či pozměňovat ani užívat způsobem či v kontextu, který by mohl působit vůči třetím osobám klamavě (data tak lze použít pouze způsobem, který nijak nemění nebo neupravuje jejich význam/smysl).

Uživatel dále bere na vědomí, že:

- otevřená data TA ČR jsou zveřejňována v souladu s právními předpisy, kterými se TA ČR řídí při výkonu své činnosti, zejména pak se zákonem č. 134/2016 Sb., o zadávání veřejných zakázek, zákonem č. 130/2002 Sb., o podpoře výzkumu, experimentálního vývoje a inovací z veřejných prostředků, zákonem č. 218/2000 Sb., o rozpočtových pravidlech a nařízení Komise EU č. 651/2014, kterým se v souladu s články 107 a 108 Smlouvy o fungování Evropské unie prohlašují určité kategorie podpory za slučitelné s vnitřním trhem,
- v případě dalšího šíření otevřených dat TA ČR je uživatel povinen vhodným způsobem uvést znění těchto podmínek užití, a to nejlépe formou přímého odkazu na tuto webovou stránku,
- tyto podmínky se nevztahují na dokumenty, které jsou chráněny autorským právem či jiným obdobným právem duševního vlastnictví třetích stran,



- otevřená data TA ČR mohou obsahovat údaje poskytnuté TA ČR třetím subjektem nebo mohou být vytvořena zpracováním údajů poskytnutých TA ČR třetím subjektem, TA ČR žádným způsobem neodpovídá za správnost, úplnost a pravdivost údajů, která byla TA ČR třetím subjektem poskytnuta a nemá povinnost správnost, úplnost a pravdivost předávaných údajů posuzovat,
- otevřená data TA ČR mají informativní povahu; TA ČR neodpovídá za případné škody vzniklé dalším užitím a zpracováním otevřených dat TA ČR jakož i za škody způsobené případnými chybami, které by otevřená data TA ČR mohla obsahovat,
- TA ČR si vyhrazuje právo z provozních, odborných či jiných důvodů dočasně omezit přístup k zveřejňovaným otevřeným datům TA ČR, a to i bez předchozího upozornění.



Příloha 2 – Text směrnice SME-38 Publikace a katalogizace otevřených dat v1

Obecná část

1. Všechny obsazené role a pozice jsou myšleny jak v mužském, tak v ženském rodě.
2. Otevírání dat je kontinuální proces. V první fázi dochází k analýze datových sad a k rozhodnutí o jejich zveřejnění. Výsledkem této fáze je publikační plán. V dalším kroku dochází k samotnému zveřejňování datových sad. To zahrnuje návrh technické podoby datových sad (formáty a datová schémata), jejich publikaci vna internetu a katalogizaci. Technologická agentura ČR provozuje svůj LKOD registrovaný v NKOD.
3. Kromě hlavních rolí vyjmenovaných v kapitole 4.1. této směrnice figurují v procesu S06 Otevřená data i další role:
 - IT specialista, který pomáhá kurátorům dat při přípravě datových sad ke zveřejnění a správci katalogu otevřených dat s provozem datového katalogu;
 - ředitel sekce interní podpory, který spouští proces otevírání dat svým rozhodnutím a obsazuje pracovníky organizace do jednotlivých rolí;
 - ředitel Kanceláře TA ČR, který schvaluje publikační plán.

Činnosti otevírání dat

1. Práce s publikačním plánem a harmonogramem zveřejňování dat probíhají výhradně pomocí aplikace ServiceDesk s použitím [sdílené tabulky publikačního plánu](#). Zjednodušený proces zveřejňování otevřených datových sad spočívá v následujících činnostech:
 - stanovení technické podoby publikovaných dat (tzv. distribucí)
 - příprava/úpravy SW obsahující datové sady na export dat
 - stanovení podmínek užití
 - export/úpravy dat v otevřeném formátu
 - vysvětlení struktury exportovaných dat (tzv. metadat) a popis vlastní distribuce
 - fyzické umístění datové sady do vhodného úložiště, příp. začlenění datové sady do lokálního katalogu otevřených dat
 - vytvoření katalogizačního záznamu v LKOD a/nebo NKOD

Popis procesu

S0601 Příprava (pravidelná revize) a zveřejnění publikačního plánu

1. Koordinátor otevírání dat rozešle jednotlivým kurátorům dat seznam doporučených datových sad uvedený ve [vzorovém publikačním plánu](#). Jednotliví kurátoři dat vyberou datové sady, které spadají do jejich kompetencí a je možné je zveřejnit v podobě otevřených dat, a případně seznam rozšíří o další datové sady v jejich kompetenci.
2. Kurátor dat vybere, dle vzorového publikačního plánu, příslušné datové sady, případně doplní publikační plán o nové datové sady vyplývající z prováděných agend.



3. Koordinátor otevírání dat ve spolupráci s kurátorem dat kontroluje, že pro každou datovou sadu v seznamu jsou doplněny všechny atributy.
4. Koordinátor otevírání dat s roční periodicitou ověřuje pravdivost a relevantnost informací obsažených v publikačním plánu (zejména ve vztahu k LKOD/NKOD).
5. V případě zjištěného nesouladu (datová sada je zastaralá, publikované datové sady se překrývají) nebo identifikace nové datové sady zajistí koordinátor otevírání dat ve spolupráci s kurátorem dat aktualizaci publikačního plánu.
6. Koordinátor otevírání dat předkládá publikační plán nebo jeho aktualizaci ke schválení ředitelům sekcí.
7. Ředitelé sekcí předložený či aktualizovaný publikační plán schválí, případně vrátí k dopracování koordinátorovi otevírání dat.
8. Koordinátor otevírání dat zabezpečí zveřejnění publikačního plánu (či jeho veřejných částí) na webových stránkách TA ČR.

S0602 Návrh na zveřejnění nové datové sady

1. Kurátor dat ve spolupráci s IT specialistou vyberou datovou sadu, vytvoří popis datové sady spolu s odůvodněním zveřejnění ve formě otevřených dat, a to prostřednictvím [Návrhu na zveřejnění dat \(formulář F-382\)](#), který se zaregistruje do aplikace ServiceDesk a dojde k notifikování koordinátora otevírání dat.

S0603 Provedení analýzy přínosů a rizik otevření datové sady

1. Koordinátor otevírání dat posoudí obdržené informace, a to obsah návrhu na zveřejnění dat a jeho odůvodnění, a informace doplní do publikačního plánu.
2. Pokud je návrh na zveřejnění dat odpovídá informacím uvedeným v publikačním plánu, předá koordinátor otevírání dat záznam analytickému oddělení k doplnění vyjádření.
3. Analytické oddělení vypracuje stanovisko k navržené datové sadě ke zveřejnění do 10 pracovních dnů a předá jej koordinátorovi otevírání dat.
4. Koordinátor otevírání dat zapracuje případné požadavky na změny od analytického oddělení a předá záznam právnímu oddělení a bezpečnostnímu řediteli k vypracování stanoviska.
5. Právní oddělení vypracuje do 10 pracovních dnů stanovisko k datovým sadám, zda je z pohledu legislativy možné danou datovou sadu zveřejnit, případně s jakými riziky, a toto stanovisko předá koordinátorovi otevírání dat.
6. Bezpečnostní ředitel vypracuje do 10 pracovních dnů stanovisko k jednotlivým datovým sadám, kde zhodnotí bezpečnostní rizika zveřejnění a rizika z pohledu ochrany osobních údajů. Toto stanovisko předá koordinátorovi otevírání dat i s případnými návrhy či doporučeními s dopadem na úpravu datových sad tak, aby bylo možné je zveřejnit (anonymizace, pseudonymizace).
7. Koordinátor otevírání dat ověří, zda právní oddělení a bezpečnostní ředitel schválili data ke zveřejnění.
8. Pokud právní oddělení či bezpečnostní ředitel data ke zveřejnění neschválili, koordinátor otevírání dat informuje kurátora dat o závažných skutečnostech, kvůli kterým nemůžou být data zveřejněna.
9. Pokud právní oddělení a bezpečnostní ředitel schválili návrh datové sady ke zveřejnění, koordinátor otevírání dat zapracuje požadavky na změny a předá záznam IT specialistovi.



10. IT specialista vypracuje do 10 pracovních dní stanovisko, ve kterém posoudí aktuální stav datové sady a vytvoří návrh na přípravu datové sady z pohledu pracnosti TA ČR a dodavatelů TA ČR pro zveřejnění v lokálním katalogu a následně předá záznam koordinátorovi otevírání dat.
11. Koordinátor otevírání dat z obdržených informací stanoví harmonogram publikace datových sad doplněním předpokládaného datumu zveřejnění datové sady a priority jejího zveřejnění (nejvyšší prioritu mají datové sady s nízkou pracností, nejvyššími přínosy a nejnižšími riziky).
12. Koordinátor otevírání dat předá kurátorovi dat schválený záznam o nové datové sadě ke zveřejnění.

S0604 Příprava datové sady ke zveřejnění

1. Kurátor dat ve spolupráci s IT specialistou zabezpečí technické zajištění zveřejnění sady dat.
2. Kurátor dat předá nově vytvořenou datovou sadu spolu s metadaty koordinátorovi otevírání dat ke kontrole.

S0605 Kontrola a publikace dat ke zveřejnění

1. Koordinátor otevírání dat zkontroluje náležitosti datové sady ke zveřejnění (data a metadata).
2. Pokud koordinátor otevírání dat zjistí v datové sadě nedostatky, předá záznam o nové datové sadě zpět kurátorovi dat. Kurátor dat nedostatky odstraní a předá opravenou datovou sadu spolu s metadaty zpět koordinátorovi otevírání dat.
3. Koordinátor otevírání dat následně zaregistruje novou datovou sadu v LKOD nebo doplní dodatečné informace k datové sadě v LKOD nebo zajistí publikaci aktualizovanou datovou sadu v LKOD.