

Automatizace kódování zdravotní péče – standardy, požadavky a řešení

Standardy vykazování v ČR

V ČR se pro vykazování zdravotní péče používají **mezinárodní i národní standardy**: diagnózy se kódují podle **MKN-10**, výkony podle **Seznamu zdravotních výkonů s bodovými hodnotami**, a pro úhradu hospitalizační péče slouží systém **CZ-DRG**. Vyúčtování péče se zdravotním pojišťovnám probíhá elektronicky formou **dávek dokladů** (např. soubory typu *K-dávka* obsahující ambulantní výkony) podle jednotné metodiky VZP.

Technické požadavky API

Moderní **REST API** pro zdravotnická data obvykle využívají standardizované formáty (např. **HL7 FHIR** s JSON/XML) a kladou důraz na **bezpečnost**. Samozřejmostí je **šifrovaná komunikace (HTTPS)** a **silná autentizace** (např. OAuth 2.0 pro přístup k citlivým údajům). Důležité je také dodržovat zdravotnické **interoperabilní standardy** pro snadnou integraci.

Existující řešení

Pro automatizaci kódování a kontrolu vykazování existují jak **open-source** nástroje, tak **komerční** produkty. Např. VZP poskytuje kontrolní software pro validaci dávek, UZIS/MZ ČR má **DRG grouper** pro klasifikaci hospitalizačních případů, a nově se objevují **AI asistenti kódování** (např. projekt *ICZ Asistent vykazování AV(D)* využívající umělou inteligenci k doporučení diagnóz a kódů výkonů). Také mezinárodní řešení jako **3M**, **Nym** apod. nabízejí automatizované kódování s využitím NLP.

Právní a etické aspekty

Zdravotnická data jsou vysoce citlivá a jejich zpracování podléhá přísným požadavkům. **GDPR** (EU 2016/679) je v ČR doplněno zákonem č. 110/2019 Sb. a klade důraz na **ochranu osobních údajů** – zdravotní údaje jsou zvláštní kategorie vyžadující např. šifrování a minimalizaci přístupu. Zákon č. 372/2011 Sb. (o zdravotních službách) zase striktně vymezuje, kdo smí nahlížet do zdravotnické dokumentace, aby bylo chráněno soukromí pacienta. Je nutné zajistit **souhlasné právní tituly** pro zpracování dat (nebo oporu v zákoně) a dodržovat lékařské tajemství.

Vstupní data AIS a zpracování

Ambulantní informační systémy (AIS) běžně generují **strukturované exporty** pro pojišťovny – typicky textové soubory v definovaném formátu (např. *soubor KDAVKA.111 pro VZP*). Tyto soubory mají přesně stanovenou strukturu (ASCII, pevné délky či separátory) a obsahují hlavičky dávek, doklady péče, kódy diagnóz, výkony atd. Efektivní parsování spočívá ve **využití známé struktury** (podle metodiky VZP) k extrakci položek nebo použití knihoven pro daný formát. Pokud jsou vstupem **klinické události v textové podobě**, je potřeba nasadit **NLP** nástroje pro analýzu textu a rozpoznání relevantních pojmů (diagnóz, výkonů) k přiřazení kódů.

Standardy vykazování zdravotní péče v ČR

Vykazovací kódy a klasifikace: Český systém vykazování zdravotní péče kombinuje mezinárodní klasifikace s národními číselníky:

- **Diagnózy** se kódují podle mezinárodní klasifikace **MKN-10** (10. revize Mezinárodní klasifikace nemocí). V ČR je tato klasifikace závazná a spravuje ji ÚZIS; česká verze MKN-10 se pravidelně aktualizuje, ale zachovává kompatibilitu s mezinárodním standardem¹. To zajišťuje jednotné kódování všech nemocí a zdravotních stavů.
- **Výkony** (lékařské úkony) se vykazují kódy ze **Seznamu zdravotních výkonů s bodovými hodnotami**. Tento seznam je vydáván vyhláškou Ministerstva zdravotnictví (počínaje vyhláškou č. 134/1998 Sb.) a obsahuje každou zdravotní službu (vyšetření, zákrok apod.) s unikátním kódem a přiřazenou bodovou hodnotou. Bodové hodnoty slouží k výpočtu úhrady. Seznam je průběžně upravován; VZP ČR publikuje jeho aktualizace formou číselníků – např. k 1.10.2025 platila verze číselníku 1453.
- **Léčiva a materiál** mají také své číselníky: např. kódy léčivých přípravků (podle Státního ústavu pro kontrolu léčiv – tzv. číselník SÚKL) a **ZUM/ZULP** kódy pro zvlášť účtovaný materiál a léčiva pro pojišťovny^{2 3}.

DRG systém: Pro lůžkovou (hospitalizační) péči se v ČR používá **DRG (Diagnosis Related Groups)** – konkrétně od roku 2020 národní varianta **CZ-DRG**. DRG klasifikuje hospitalizační případy do skupin na základě diagnóz, provedených výkonů a dalších faktorů, s cílem ohodnotit ekonomickou náročnost péče jednotně. Tento systém umožňuje srovnat klinicky podobné hospitalizace z hlediska nákladů a určovat paušální úhradu za případ. **CZ-DRG** je spravováno ÚZIS a průběžně aktualizováno (ČSÚ vydává sdělení o aktualizaci klasifikace CZ-DRG v návaznosti na §19 zákona č. 89/1995 Sb.). Pro praktické využití existují tzv. **groupery** – software, který na vstupu přijme kódy diagnóz a výkonů pacienta a přiřadí mu DRG skupinu. Např. Ministerstvo zdravotnictví poskytuje oficiální grouper CZ-DRG (aktuálně ve verzi 3.0/6.0) a komerční variantu grouperu nabízí i společnosti jako Asseco CE pro pojišťovny a nemocnice.

Struktura a formát dávek pro pojišťovny: Vykazování vůči zdravotním pojišťovnám probíhá **dávkově**. Poskytovatelé (lékaři, nemocnice) v pravidelných intervalech (obvykle měsíčně) odesílají pojišťovně **dávku dokladů** – elektronický soubor obsahující více záznamů o provedené péči. Pro tento účel existuje jednotný **Datový rozhraní** definované zejména VZP (které jako největší pojišťovna udává standard akceptovaný i ostatními ZP).

¹<https://mkn-kodovani.cz/avd-ii/>

²<https://mkn-kodovani.cz/avd-ii/>

³<https://mkn-kodovani.cz/avd-ii/>

- Nejrozšířenějším formátem je textový **soubor KDAVKA.XXX** (kde XXX je kód pojišťovny – pro VZP 111)^{4 5}. Jde o ASCII soubor v kódování Latin2 s pevně stanovenou strukturou. Soubor může obsahovat jednu nebo více **dávek**. Každá **dávka** má úvodní řádek (záhlaví dávky) určující typ dávky a základní údaje (např. identifikace poskytovatele, období, počet dokladů), za kterým následují jednotlivé **doklady** péče. **Doklad** typicky představuje vyúčtování péče pro jednoho pacienta za 1 návštěvu či hospitalizaci a obsahuje více vět: hlavičku dokladu (s údaji o pacientovi, diagnózách atd.) a následně detailní řádky výkonů, materiálů, léků apod. podle potřeby.
- **Druhy dokladů a typy dávek** jsou přesně definovány číselníky VZP^{6 7}. Například ambulantní péče se vykazuje dokladem typu 01 (Vyúčtování výkonů v ambulantní péči) a tyto doklady se mohou sdružovat do *ambulantní dávky*. „**K-dávka**“ v praxi obvykle označuje právě dávku obsahující ambulantní výkony. Podle datového rozhraní VZP se pro ambulantní smíšenou dávku používá **typ 98** (Dávka ambulantní smíšená) – ta obsahuje doklady druhu 01 (ambulantní výkony) a případně 03 (zvlášť účtované materiály)⁸. Soubory KDAVKA.111 tedy většinou obsahují dávku typu 98 pro dané období, případně více dávek za různá odbornostní či administrativní členění.

Validace vykazovacích kódů: Pojišťovny při příjmu dávek provádějí **automatické kontroly**. Dávka prochází sadu validačních pravidel a pokud **chyba** poruší pravidla, může pojišťovna **odmítnout** buď celou dávku, nebo konkrétní doklad. Mezi typické kontrolované aspekty patří např.:

- Platnost kódů a existence v číselnících ke dni poskytnutí péče (např. zda použitý kód výkonu či diagnózy je platný).
- Kombinace výkonů a diagnóz – některé výkony lze vykázat jen u určitých diagnóz či odborností lékaře, jiné kombinace jsou nepřípustné.
- Frekvenční omezení – některé výkony lze vykázat jen omezeně (např. určité preventivní prohlídky jednou za rok apod.), systém ověřuje, zda limit nebyl překročen.

⁴<https://media.vzpstatic.cz/media/Default/dokumenty/vyuctovani/verze-metodiky-a-datoveho-rozhrani/metodika-vyuctovani-datove-rozhrani-individualnich-dokladu.pdf>

⁵<https://media.vzpstatic.cz/media/Default/dokumenty/vyuctovani/verze-metodiky-a-datoveho-rozhrani/metodika-vyuctovani-datove-rozhrani-individualnich-dokladu.pdf>

⁶<https://www.vzp.cz/poskytovatele/ciselniky/druh-dokladu-a-tyt-davky-vzp>

⁷<https://www.vzp.cz/poskytovatele/ciselniky/druh-dokladu-a-tyt-davky-vzp>

⁸<https://www.vzp.cz/poskytovatele/ciselniky/druh-dokladu-a-tyt-davky-vzp>

- Úplnost dokladu – například zda u hospitalizace jsou uvedeny potřebné kódy (hlavní a vedlejší diagnózy, DRG markery atd.), nebo zda nechybí kód indikující lékaře, IČP pracoviště apod.

Metodika VZP detailně popisuje tyto kontroly. VZP navíc poskytuje smluvním partnerům program **VZP Kontrol**, který umožňuje zkontrolovat dávky před odesláním. Pokud se v dávce najdou chyby, poskytovatel je může opravit před oficiálním podáním, čímž se vyhne odmítnutí. **Nejčastější chyby** ve vykazování zahrnují např. nesoulad rodného čísla a kódu pojišťovny, duplicity dokladů, chybné pořadí výkonů, opomenuté vykazání některého souvisejícího úkonu atd.⁹. Systém pojišťovny pak k chybné dávce vystaví tzv. **validační protokol** s výpisem zjištěných chyb (tzv. *vratky*), které se zasílají zpět poskytovateli k nápravě¹⁰.

Shrnutí: Standardy vykazování v ČR jsou robustní a jednotné – využívá se MKN-10 pro diagnózy, národní číselník výkonů pro procedury, DRG pro hospitalizace a jednotné datové rozhraní pro komunikaci s pojišťovnami (dávky dokladů). Důsledná validace kódů je klíčová pro správné proplacení péče, což je právě jedním z cílů uvažovaného projektu (automatická kontrola a návrh kódů před vykazáním).

Technické požadavky na REST API pro zdravotnická data

Návrh REST API služby pro zpracování zdravotnických dat musí respektovat jak obecné principy robustních API, tak **specifické požadavky zdravotnictví** na formát dat a zabezpečení:

- **Datové formáty a standardy:** V moderním zdravotnictví se prosazuje standard **HL7 FHIR (Fast Healthcare Interoperability Resources)** pro výměnu dat přes REST API. FHIR poskytuje jednotné datové modely (Resources) pro pacienty, návštěvy, diagnózy, výkony, pojištění apod., které lze přenášet ve formátu JSON nebo XML přes HTTP. **FHIR je navržen pro webové API** – definuje RESTful operace pro čtení, vyhledávání, zápis dat apod.¹¹. Použití FHIR by tedy umožnilo kompatibilitu s dalšími systémy (např. nemocničními informačními systémy třetích stran) a snadné rozšíření služby o nové typy zdravotnických dat. Pokud by však integrace FHIR byla příliš komplexní, lze zvolit proprietární JSON/XML formát, avšak **doporučuje se držet zavedených standardů** pro budoucí interoperabilitu. Například platformy jako InterSystems IRIS či Microsoft Azure Health používají FHIR jako základ pro API nad

⁹<https://mkn-kodovani.cz/avd-ii/>

¹⁰<https://mkn-kodovani.cz/avd-ii/>

¹¹<https://mkn-kodovani.cz/avd-ii/>

zdravotnickými daty¹², což potvrzuje, že FHIR je považován za „*budoucnost interoperability ve zdravotnictví*“¹³.

- **Tradiční formáty AIS:** Je dobré zmínit, že současné AIS v ČR často komunikují se světem pomocí starších standardů jako je **DASTA** (národní formát pro výměnu zdravotních dat) nebo pomocí specifických souborů (jako právě zmíněné dávky pro pojišťovny). Tyto formáty nejsou REST API, ale soubory či zprávy. V kontextu navrhované služby by tedy API mohlo přijímat i **soubory K-dávky** jako jeden z typů vstupu (např. upload textového souboru s dávkou k validaci). Pro parsing takového souboru může serverová aplikace využít buď vlastní parser podle dokumentace VZP, nebo otevřené knihovny, pokud existují. Případně by bylo možné převést nahraný soubor do interní struktury – například konvertovat jej do interního FHIR Bundlu (kde každý doklad = FHIR Claim resource, diagnózy = Condition, výkony = Procedure nebo Observation apod.). Tím by se zjednodušilo následné zpracování pravidly.
- **Bezpečnost přenosu (Transport Layer Security):** Veškerá komunikace s API musí probíhat **šifrovaně**. Základním požadavkem je použití **HTTPS** pro všechny endpointy, čímž se ochrání citlivá data pacientů při přenosu před odposlechem či úpravou. Samozřejmostí je použití aktuálních verzí protokolu TLS a pravidelná údržba certifikátů.
- **Autentizace a autorizace:** Zdravotnická data spadají pod **GDPR zvláštní kategorie údajů**, takže přístup k API musí být přísně kontrolován. **OAuth 2.0** je dnes standardem pro zabezpečení REST API – umožňuje vydávat přístupové tokeny klientům a definovat rozsah (scopes) jejich oprávnění. Například lze použít **OAuth2.0 Client Credentials** flow pro serverového klienta (pokud by službu volal přímo informační systém nemocnice), nebo **OAuth2 Authorization Code** flow s OpenID Connect, pokud by přístup byl i pro uživatele. Ověření pomocí OAuth 2.0 je považováno za osvědčený způsob – např. i implementace FHIR od Oracle používá OAuth 2.0 k zabezpečení API¹⁴. Alternativou či doplňkem mohou být **certifikáty** (mutuální TLS) u důvěryhodných integrací – v ČR například komunikace s pojišťovnami v režimu B2B vyžaduje klientský certifikát. V rámci interního API bychom však volili spíše OAuth2/JWT, protože je flexibilnější pro různé klienty.
- **Oprávnění a audit:** Každý požadavek by měl být **autorizován** tak, aby klient měl přístup jen k datům, která smí vidět (např. jen data vlastní organizace). Zároveň je vhodné implementovat **audit logging** – zaznamenávat kdo a kdy přistoupil k jakým datům, což je požadavek plynoucí jak z dobré praxe, tak z případných legislativních

¹²<https://mkn-kodovani.cz/avd-ii/>

¹³<https://mkn-kodovani.cz/avd-ii/>

¹⁴<https://docs.oracle.com/en/industries/health-sciences/healthcare-data-repository/8.2.0.3/fhir-guide/working-fhir-rest-apis.html>

nároků (např. zákon o zdravotních službách vyžaduje, aby poskytovatel vedl záznamy o přístupech k elektronické dokumentaci).

- **Formát vstupů a výstupů:** API by mělo přijímat vstupy v dohodnutém formátu. Pravděpodobně:
 - Pro **user story 1** (validace kódů existujících dokladů) – vstupem může být buď celý soubor dávky (např. jako příloha požadavku) nebo datová struktura odpovídající jednomu dokladu/dávce v JSON. Výstupem pak podrobný report chyb a varování pro každý doklad, případně návrhy úprav.
 - Pro **user story 2** (generování kódů z klinické události) – vstupem může být text (např. ve formě strukturovaného JSON: obsahující anamnézu, závěr, provedené výkony slovně) nebo už předstrukturovaná data (např. klíč: hlavní diagnóza textem, seznam problémů, provedené zákroky popisem). Výstupem by byl seznam navržených kódů (diagnóz, výkonů, případně materiálů) ideálně i s **odůvodněním** – např. párování k úryvkům vstupního textu („*kód XYZ přiřazen na základě zmínky o ...*“). Formát výstupu by opět byl JSON obsahující navržené kódy a meta informace.
- **Výkon a škálovatelnost:** Zpracování může být náročné (zejména generování kódů pomocí AI/NLP). API by mělo být navrženo s ohledem na škálovatelnost – např. asynchronní zpracování pro větší dávky (klient zadá požadavek a API vrátí ID úlohy, která se zpracuje na pozadí, a klient si později stáhne výsledek). Také je nutné počítat s **odolností** – validace i návrh kódování musí probíhat s maximální spolehlivostí, aby se nestalo, že chyba služby zdrží vykazování péče.
- **Další aspekty:** Vzhledem k citlivosti dat je důležité také **logování a monitoring** provádět opatrně (nejlépe logovat jen technické informace, ne samotná zdravotní data, nebo logy adekvátně zabezpečit/anonymizovat). Při návrhu API je vhodné začlenit i **verzování** – zdravotnické standardy se vyvíjejí (např. změny v číselnících, nové DRG verze), takže API by mělo umožnit verzovat endpointy či parametry, aby změny nerozbily integrace.

Shrnutí: Technologicky by služba měla využít moderní standardy – **FHIR/JSON pro strukturu dat** a ověřené postupy zabezpečení (HTTPS, OAuth2). Tím zajistí interoperabilitu a důvěryhodnost při zpracování citlivých zdravotnických informací. Stávající formáty (K-dávky, DASTA) by měla umět převzít a převést, aby zapadla do ekosystému českých AIS.

Existující řešení pro automatizaci kódování a vykazování

Před zahájením vývoje je užitečné znát, jaká řešení již na trhu existují či jsou ve vývoji pro podobné účely – ať už volně dostupná (open-source) nebo komerční. Níže uvádíme přehled několika takových řešení:

- **Kontrolní a vykazovací software zdravotních pojišťoven:** V ČR tradičně každá větší zdravotní pojišťovna poskytuje zdravotnickým zařízením nástroje pro podání a

kontrolu dávek. Například **VZP Kontrol** (interní nástroj VZP) umožňuje prověřit dávky před odesláním podle aktuální metodiky VZP. Takový software odhalí formální chyby a některé věcné nesrovnalosti (např. špatné kombinace kódů) a vygeneruje report. Je typicky poskytován zdarma smluvním partnerům, ale není open-source a funguje spíše jako samostatná aplikace než jako API. Přesto představuje **státem uznávaný etalon validace** (díky vazbě na metodiky VZP).

- **DRG Groupery:** Pro přiřazení DRG k hospitalizačním případům existují nástroje jako **CZ-DRG Grouper** od ÚZIS/MZ ČR (dostupný pro nemocnice i pojišťovny). Komerční variantu má např. **Asseco DRG Grouper**, který umí velmi rychle zpracovat velké objemy dat a integrovat se do nemocničních systémů. Tyto groupery ale řeší jen část problému – **klasifikaci hospitalizačních případů** – a často předpokládají, že diagnózy a výkony už jsou správně zakódované. Do našeho scénáře by grouper zapadal spíše jako následný krok (např. po přiřazení kódů udělat i DRG zařazení případů).
- **Nástroje pro podporu kódování diagnóz a výkonů pomocí AI:** Zajímavou novinkou je prototyp od společnosti ICZ a ZČU Plzeň s názvem **AV(D) – Asistent vykazování**. Jde o inteligentního asistenta, který pomocí umělé inteligence dokáže na základě dat o pacientovi a průběhu léčby **doporučit lékaři vhodné kódy**: navrhne hlavní a vedlejší diagnózy, relevantní výkony, léky či zdravotnické prostředky k vykazování^{15 16}. Projekt AV(D) má módy „*porad' mi*“ (návrh kódů) a „*zkontroluj*“ (validace již zadaných kódů)¹⁷, což se výrazně překrývá s funkcemi našeho zamýšleného API. Zatím jde o pilotní projekt (výstupy byly prezentovány v roce 2023 a 2024), ale ukazuje trend nasazení **NLP a machine learningu** v této doméně. Nástroj AV(D) není veřejně dostupný ani open-source, jde o proprietární vývoj.
- **Mezinárodní komerční řešení pro Computer-Assisted Coding (CAC):** Ve světě (zejm. v USA) existují software pro automatizované kódování zdravotní péče z dokumentace. Například **3M 360 Encompass** a podobné produkty automatizují proces kódování tím, že analyzují elektronické dokumenty pacienta (propouštěcí zprávy, lékařské poznámky) a navrhují kódy ICD-10, CPT atd. Společnost **Nym Health** vyvinula AI engine, který **autonomně kóduje** ambulantní i nemocniční zprávy s využitím zpracování přirozeného jazyka, a chlubí se vysokou přesností bez zásahu

¹⁵<https://mkn-kodovani.cz/avd-ii/>

¹⁶<https://mkn-kodovani.cz/avd-ii/>

¹⁷<https://mkn-kodovani.cz/avd-ii/>

člověka^{18 19}. Taková řešení jsou většinou komerční (licencovaná software-as-a-service), integrují se do nemocničních informačních systémů a cílí na zrychlení revenue cycle managementu. Jejich **výhoda** je dlouholetý vývoj a ověřené algoritmy, nevýhodou pak cena a nutnost přizpůsobení lokálním kódovacím pravidlům (americké kódovací standardy jako CPT/HCPCS se liší od českých).

- **Open-source projekty:** V oblasti **validace vykazování** nejsou příliš známy open-source projekty pokrývající českou metodiku (ta je dost specifická). Nicméně dílčí open-source nástroje existují: například pro DRG lze uvést knihovnu **DRGPy** (Python) pro přiřazování amerických Medicare DRG, případně projekt **OpenDRG** pro čínský DRG – tyto nejsou přímo použitelné pro CZ-DRG, ale ukazují, že určitá logika DRG se dá otevřeně implementovat. Pro **NLP extrakci diagnóz** by bylo možné využít obecné modely (např. otesný jazykový model vytrénovaný na zdravotnických textech), avšak customizace pro český jazyk a terminologii by byla nutná. Z volně dostupných komponent lze zmínit např. **MIMIC-IV dataset a modely** (anglicky) nebo nástroje jako **Apache cTAKES** či **SpaCy med** – ty by ale musely být adaptovány do češtiny a na české klasifikace.

Srovnání vybraných řešení

Následuje tabulka shrnující některé existující systémy a jejich vlastnosti z hlediska funkcí, formátů a výhod či nevýhod:

Řešení	Typ	Hlavní funkce	Podporované formáty/datové vstupy	Výhody	Nevýhody
VZP Kontrol (oficiální kontrolní program)	Komerční (distribován zdarma partnerům, uzavřený)	Validace dávek podle aktuální metodiky pojišťovny (formální kontroly, věcné kontroly kódů)	Vstup: soubor <i>KDAVKA.xxx</i> (ASCII podle definice VZP) Výstup: protokol chyb (textový/PDF)	+ Aktuálnost dle metodiky VZP+ Odhalí většinu běžných chyb v kódování+ Ověřený pojišťovna mi (oficiální nástroj)	– Není API (nutnost ručního použití)– Neposkytují návrhy kódů, jen kontroly– Uzavřený software (nelze modifikovat)

¹⁸<https://media.vzpstatic.cz/media/Default/dokumenty/vyuctovani/verze-metodiky-a-datoveho-rozhrani/metodika-vyuctovani-datove-rozhrani-individualnich-dokladu.pdf>

¹⁹<https://media.vzpstatic.cz/media/Default/dokumenty/vyuctovani/verze-metodiky-a-datoveho-rozhrani/metodika-vyuctovani-datove-rozhrani-individualnich-dokladu.pdf>

CZ-DRG Grouper(UZIS/MZ ČR, Asseco)	Komerční (nemocnice, ZP)(existuje i webová verze UZIS)	Přiřazení DRG skupiny hospitalizačnímu případu podle zadaných kódů diagnóz, výkonů a dalších údajů	Vstup: sada kódů (DG, výkonů, věk pac., apod.), možno přes UI nebo datový soubor Výstup: DRG kód + související ukazatele (váha, pobyt)	+ Rychlá klasifikace i velkých objemů dat+ Oficiální algoritmus (shoda s úhradovým mechanismem)+ Lze integrovat do NIS	– Řeší pouze výstupní DRG , neověřuje validitu samotných kódů vstupu– Nevhodné pro ambulantní péči (DRG je pro lůžka)– Uzavřené jádro (není open-source)
ICZ AVD (Asistent vykazování)(výzkumný projekt)	Interní prototyp (AI asistence)	Doporučování kódů diagnóz a výkonů z textové dokumentace; Kontrola zadaných kódů vůči dokumentaci	Vstup: texty zdravotnické dokumentace (např. propouštěcí zpráva) – integruje se do klinického systému Výstup: Návrh hlavní a vedlejší DG , kódů výkonů , léčiv atd., vč. pravděpodobností ²⁰	+ Využívá umělou inteligenci k porozumění textu+ Může zvýšit úplnost vykazování (navrhne opomenuté položky)+ Funkce „porad“ i „zkontroluj“ – dvojí užití	– Ve fázi pilotu (2023/2024), není běžně dostupný– Potenciální nutnost doladění přesnosti (AI může chybovat)– Neznámé licenční podmínky (pravděpodobně placené řešení v budoucnu)
3M 360 Encompass(USA)	Komerční (pro nemocnice)	Computer-Assisted Coding: automatické kódování z dokumenta	Vstup: elektronická dokumentace pacienta (anglický text, strukturovaná	+ Ověřená technologie nasazená ve velkých nemocnicích+	– Nepodporuje české kódy z krabice (nutná

²⁰<https://mkn-kodovani.cz/avd-ii/>

		ce + podpora coderů, integrace přímé do HIM workflow	data z EHR)Výstup: Kódy (ICD-10-CM, ICD-10-PCS, CPT, DRG) – dle US standardů	Komplexní – pokrývá i audit, compliance, reporty+ Úspora času kódování, vysoká přesnost s lidským dohledem	lokalizace)– Velmi nákladné řešení (licence, implementace)– Určeno primárně pro anglické lékařské texty
Nym Health(USA AI solution)	Komerční (cloudová AI služba)	Autonomní kódování – plně automatické přiřazení kódů z lékařských zpráv pomocí NLP, bez nutnosti ruční validace	Vstup: text lékařské zprávy (ENG) nebo strukturovaná data z EHRVýstup: Sada kódů (ICD-10, CPT) připravená k vykazání, s vysvětlením	+ Pokročilé NLP trénované na klinický jazyk ²¹ + Rychlé zpracování (sekundy na dokument)+ Slibuje snížení potřeby lidských kódovačů	– Angličtina a US předpisy – nutno přetrénovat pro ČR– Uzavřená služba, data by musela být posílána do cloudu (možný problém s GDPR)– Cena a nutnost smlouvy s dodavateli

Pozn.: Kromě výše uvedených existují i další české lékařské systémy, které mají dílčí funkce pro ulehčení kódování – např. ambulantní software často kontroluje, zda lékař zadal všechny potřebné údaje pro vykazání (rodné číslo pacienta, pojišťovnu, hlavní diagnózu atd.), případně integruje online služby (např. **B2B pojišťoven** na ověření pojištěnce v registru – viz funkcionality FONS Akord pro on-line validaci čísla pojištěnce přes VZP rozhraní). Tyto funkce však většinou neprovádějí komplexní návrh kódů, spíše základní kontroly.

Z přehledu je zřejmé, že **naš projekt kombinující validaci kódů a jejich automatický návrh by v ČR byl poměrně unikátní** (podobným směrem jde jen zmíněný AV(D) prototyp). Využit bychom mohli poznatky z existujících řešení: validační pravidla odvodit z metodiky VZP (případně inspirovat se výstupy VZP Kontrol) a pro AI generování kódů se učit z přístupů,

²¹<https://media.vzpstatic.cz/media/Default/dokumenty/vyuctovani/verze-metodiky-a-datoveho-rozhrani/metodika-vyuctovani-datove-rozhrani-individualnich-dokladu.pdf>

kteřé volí AV(D) či zahraniční Nym (tzn. trénovat model na reálných datech zdravotnické dokumentace a vykázaných kódů).

Právní a etické požadavky na zpracování zdravotnických dat

Zpracování dat o pacientech je vázáno přísnými **legislativními pravidly** a etickými zásadami, které nelze opomenout při návrhu služby:

- **Ochrana osobních údajů (GDPR):** Veškeré údaje o zdravotním stavu jsou podle GDPR považovány za **zvlášť citlivé osobní údaje** (tzv. zvláštní kategorie osobních údajů dle čl. 9 GDPR). Jejich zpracování je možné pouze za speciálních podmínek – v zdravotnictví typicky na základě toho, že *“zpracování je nezbytné pro účely zdravotní péče či veřejného zdraví”* a provádí jej subjekt vázaný zákonnou povinností mlčenlivosti. Český **adaptační zákon č. 110/2019 Sb.** doplňuje GDPR a upřesňuje povinnosti správců. Pro náš případ, kdy by poskytovatelem služby byla zřejmě zdravotnická zařízení nebo jejich IT partner, by právním základem bylo plnění zákonné povinnosti (vykazování péče pojišťovněm je dané zákonem o veřejném zdravotním pojištění) – tedy není nutný individuální souhlas pacienta, ale **důraz na zabezpečení** zůstává. GDPR vyžaduje provést **analýzu rizik** a případně DPIA (posouzení vlivu na ochranu osobních údajů), zajistit minimalizaci dat (zpracovávat jen údaje nutné pro daný účel), a **uložit data jen po nezbytnou dobu**. Pokud by služba využívala cloud nebo předávala data třetí straně (např. volání AI služby), je třeba obzvlášť pečlivě ošetřit smluvně i technicky, aby nedošlo k neoprávněnému přenosu dat mimo OKOLÍ EU.
- **Zákon o zdravotních službách č. 372/2011 Sb.:** Tento zákon upravuje vedení zdravotnické dokumentace a povinnost **mlčenlivosti** zdravotnických pracovníků. Stanoví, že poskytovatel zdravotních služeb je povinen zabezpečit zdravotnickou dokumentaci před zneužitím a že do ní mohou nahlížet jen oprávněné osoby (pacient, ošetřující lékaři, orgány kontroly atd.)²². Pro naši službu to znamená, že **data pacientů nesmí být zpřístupněna nikomu neoprávněnému**. Pokud by například API provozovala externí firma, musí být smluvně **zpracovatelem** osobních údajů se všemi garancemi (dle GDPR i zákona 372/2011 Sb.). Také jakýkoli vývojář či administrátor systému musí být vázán mlčenlivostí. Zákon také určuje, jak dlouho se dokumentace uchovává a jak se likviduje – při návrhu úložiště výsledků (logy, databáze) tedy musíme zohlednit skartací lhůty (v ambulantní péči obvykle 5-10 let po posledním záznamu, u některých dat déle podle vyhlášky o zdravotnické dokumentaci).
- **Zákon o veřejném zdravotním pojištění č. 48/1997 Sb.:** Tento zákon sice primárně řeší nároky z pojištění, ale obsahuje i ustanovení o povinnosti poskytovatelů vykazovat pojišťovněm poskytnutou péči a používat při tom předepsané formy a

²²<https://www.vzp.cz/poskytovatele/vyuctovani-zdravotni-pece>

číselníky (MZ ČR má podle §17 odst.4 zmocnění vydávat vyhlášky stanovující seznam výkonů a další číselníky pro účely úhrad). Pro nás z toho plyne, že služba musí udržovat aktuálnost dle těchto předpisů (např. když vyjde nová vyhláška měnící kódy výkonů, systém to musí reflektovat). Zákon také počítá s **kontrolami ze strany pojišťoven** – výstupy naší služby by teoreticky mohly být využity i při auditu (např. log doporučení by mohl sloužit jako doklad, proč lékař vykázal určité kódy). Tato sekundární použitelnost ale hlavně znamená, že systém musí být **transparentní** (aby šlo obhájit před revizním lékařem, že doporučil správný postup kódování).

- **Etické zacházení s daty a AI:** Pokud bychom využívali umělou inteligenci k návrhu kódů, vstupujeme i do oblasti etiky AI. Je nutné zajistit, aby model nebyl **zkreslený** (např. nenavrhne systematicky nadhodnocené kódování – což by bylo neetické vůči pojišťovně, nebo naopak nepodhodnocuje, což by zas mohlo uškodit poskytovateli). Model by měl být trénován na reprezentativních datech a výstupy by měly procházet **revizí** (alespoň v pilotní fázi určitě musí reálný lékař kontrolovat, co AI doporučila). Také je třeba myslet na **informovaný souhlas** v širším smyslu: lékaři by měli vědět, že při používání takového nástroje stále nesou odpovědnost za vykazání péče a že návrh AI je jen pomůcka, nikoli závazná instrukce.
- **Lokalizace dat:** Vzhledem k legislativě (zejm. GDPR) je ideální, aby veškerá data byla zpracovávána **lokálně nebo v datových centrech v EU**. Zpracování zdravotních dat mimo EU by bylo právně komplikované (nutnost standardních smluvních doložek atd.). Proto pokud by řešení využilo např. cloud, mělo by jít o EU datacentrum nebo privátní cloud poskytovatele zdravotních služeb.

Shrnutí: Právní rámec v ČR vyžaduje, aby navržená služba měla **vestavěnou ochranu soukromí (privacy by design)**. To znamená špičkové zabezpečení, přístup jen pro autorizované subjekty, transparentnost a soulad s účelem (vykazování péče, které je zákonnou povinností). Eticky je nutné, aby systém pomáhal zlepšit péči a spravedlivě vykazovat – nikoli se stát nástrojem k nadužívání úhrad či diskriminaci. V praxi to obnáší důkladné testování, právní konzultace (zejm. DPO – pověřenec pro ochranu osobních údajů – by měl projekt posvětit) a průběžné revize souladu s předpisy, zejména při změnách legislativy.

Vstupní data z AIS a jejich parsování a analýza

Běžné typy vstupů v AIS: Ambulantní informační systémy typicky pracují s následujícími datovými formáty, které jsou relevantní pro naše účely:

- **Exporty dávek pro pojišťovny:** Jak již bylo zmíněno, AIS dokáže vygenerovat soubor (např. *KDAVKA.111*) s vyúčtováním za určité období. Tento soubor je strukturovaný podle datového rozhraní. Například začíná hlavičkou dávky (s kódem typu dávky, IČZ/IČP zařízení, období, počtem dokladů), pak následuje každý doklad – ten sám může mít víc řádků (věty 1. typu – základ dokladu, 2. typu – položky dokladu). Formát je často **fixed-width text** (každá položka má předdefinované pozice v řádce) nebo

oddělený speciálními znaky. **Parsování** takového vstupu vyžaduje mít k dispozici specifikaci: např. vědět, že znaky 1–2 představují druh dokladu, 3–5 pořadí dokladu v dávce, 6–15 rodné číslo, 16–20 kód diagnózy atd. – toto je dáno dokumentací VZP. Existují implementace uvnitř AIS, ale pro vlastní službu bychom museli buď implementovat parser ručně dle dokumentace, nebo použít nějakou knihovnu (pokud by existovala pro zpracování VZP dávek). Vzhledem k rozšířenosti by mohla existovat knihovna třeba v C# nebo Java (používaná v nemocničních systémech), ale není veřejně známá open-source varianta, takže spíše půjde o vlastní vývoj. Výstupním strukturovaným objektem pak bude např.: [{doklad1: {pojistovna: 111, diag: ["J12.3"], vykony: ["17101", "09513"...], ...}}, {...}] – tedy datová struktura vhodná k dalšímu zpracování.

- **DASTA a jiné exporty:** Některé starší AIS používají **DASTA** standard pro obecné exporty zdravotních dat (typicky laboratorní výsledky, epikrízy apod.). DASTA je textový formát, využívající segmenty podobně jako HL7 v2.x (s delimitem |). Pro naše účely by DASTA mohla být relevantní, pokud by AIS uměl vyexportovat celou zdravotní dokumentaci pacienta nebo události tímto způsobem. Např. by mohla existovat DASTA věta obsahující *propouštěcí zprávu*, z níž by náš systém vytěžil kódy. DASTA parsování má k dispozici definici od ČMIS, a na trhu jsou komponenty (např. firma STAPRO měla modul na DASTA komunikaci). Pokud by však AIS přímo poskytl **HTML nebo PDF zprávy**, museli bychom nasadit krok OCR/extrakce textu z PDF a pak NLP analýzu – to je technicky možné, ale znamenalo by to větší složitost (raději chceme strukturovanější data, pokud možno).
- **Interní databázové struktury AIS:** V ideálním případě by integrace proběhla tak, že AIS předá přímo **strukturovaná data** – např. voláním našeho API s JSON objektem: { pacient: {id, pojistovna}, diagnózy: [...], ukony: [...], textZpravy: "..."} . To by bypassovalo nutnost parsovat textové exporty. Realita je však taková, že přímý přístup k interním DB AIS třetí stranou bývá omezený. Proto se nejspíš budeme opírat o oficiální *exportní formáty* (dávky, zprávy).

Efektivní parsování a analýza:

- **S využitím definic formátu:** Pro strukturované vstupy (jako je K-dávka) je nejefektivnější psát parser "natažený" přesně na formát. Tedy např. načíst soubor řádek po řádku a podle pozice v souboru ukládat hodnoty do objektu. Díky tomu, že formát je stabilní (mění se jen s verzemi datového rozhraní – které vychází třeba jednou ročně), můžeme parser udržovat relativně snadno. Důležité je také validovat **syntaktickou správnost** – např. spočítat kontrolní součty, pokud jsou ve formátu (někdy dávka má údaj o počtu dokladů a součtu bodů, který musí sedět s obsahem). To pomůže odhalit, zda při přenosu nedošlo k poškození dat.
- **Nástroje pro parsing:** V případě CSV/TSV výstupů (některé AIS malé ambulance umožňují export do CSV) lze použít standardní CSV parser. U fixed-width textů je možné využít např. Python s modul struct nebo pandas (definovat widths) apod. U

HL7 v2 existují knihovny (hl7apy v Pythonu, NHapi v .NET). Pro FHIR formát by byly k dispozici hotové parsery JSON stačí standardní (protože FHIR používá JSON schema).

- **Analýza textových vstupů (NLP):** Pokud vstupem není předem zakódovaný doklad, ale **klinická textová data** (např. lékařská zpráva, poznámka z ambulance), pak je třeba nasadit zpracování přirozeného jazyka. Pro češtinu to představuje výzvu, ale dá se postupovat takto:
 - a. **Předzpracování textu:** odstranit šum (např. čísla stránek, duplicity), rozdělit text na věty, tokenizovat slova. V češtině je potřeba i **lematizace** či ohled na skloňování, protože např. diagnóza může být uvedena ve tvaru „Frakturou femuru“ a my chceme rozpoznat, že jde o pojem „fraktura femur“.
 - b. **Rozpoznání klíčových medicínských pojmů:** to lze dělat buď pravidlově – mít slovník diagnóz a výkonů. ÚZIS poskytuje číselník MKN-10 v elektronické formě²³, takže můžeme mít seznam všech názvů diagnóz a zkusit je najít v textu (např. regulárně). Podobně pro výkony existují popisy výkonů, ale ty jsou dlouhé texty – spíše by se hledaly klíčová slova typu „vyšetření“, „sono“, „operace slepého střeva“ atd. Lepší přístup je využít **modely strojového učení**: natrénovat pojmenované entity (NER) pro lékařské koncepty. Existují předtrénované modely např. pro angličtinu (UMls encyklopedie), pro češtinu bychom museli trénovat třeba na upravených datech nebo použít obecný multijazyčný transformer (XLM-Roberta, mBERT) doladěný na anotovaných zdravotnických textech. To je však náročnější část projektu.
 - c. **Mapping na kódy:** jakmile máme z textu rozpoznané, že např. obsahuje pojem „hypertenze“ nebo synonymum „vysoký krevní tlak“, přiřadíme tomu odpovídající kód diagnózy (I10 pro esenciální hypertenzi v MKN-10). To vyžaduje mít **mapování lékařských termínů na kódy**. Pro diagnózy to lze udělat přes slovník (NZIP třeba nabízí elektronický číselník MKN-10²⁴). Pro výkony je to složitější – názvy výkonů jsou dost technické, ale šlo by to (např. „Sonografie břicha“ -> výkon 89213). Případně by se musel využít expertní znalost zakódovaná do pravidel nebo opět ML model, který rovnou predikuje kódy.
 - d. **Validace navržených kódů:** poslední krok – než je výsledek vrácen, měl by projít stejnými validačními pravidly jako v user story 1. Tedy pokud NLP navrhne kombinaci kódů, která nesmí být spolu, tak to ještě upravit nebo označit jako nejisté.
- **Výkon při zpracování batch vs single:** Parsování strukturovaných vstupů (dávek) je rychlé i pro stovky dokladů (textový soubor o pár MB se parseje v řádu sekund). Naproti tomu NLP analýza textu může trvat déle (desítky sekund na dokument u

²³<https://mkn-kodovani.cz/avd-ii/>

²⁴<https://mkn-kodovani.cz/avd-ii/>

velkých modelů, nebo nutnost zajistit GPU akceleraci pro rychlejší zpracování). Možné řešení je ke službě přistupovat různě dle režimu: pro **validaci dávků** (může mít stovky dokladů) se použije rychlý deterministický modul (pravidla), pro **návrh kódů z textu** se to dělá po jednom dokumentu a může to interně použít frontu úloh a paralelizaci.

Shrnutí: Typické vstupy z AIS – ať už ve formě oficiálních exportů nebo surových klinických záznamů – lze zpracovat, pokud známe jejich formát. Pro dávky pojišťoven využijeme fixní strukturu (jednoznačné mapování polí), pro volnější klinická data nasadíme kombinaci slovníků a AI. Klíčové je, že **naš systém musí udržovat krok s formáty AIS**: kdyby se např. v budoucnu přešlo na XML dávky nebo FHIR komunikaci (což nelze vyloučit, MZČR do budoucna směřuje k využití evropských standardů), služba by se tomu měla přizpůsobit. Z hlediska implementace tedy počítáme s **modularitou** – mít modul pro zpracování VZP textových dávek, modul pro NLP z volného textu, případně modul pro FHIR, pokud by některý AIS uměl poslat data přímo jako FHIR Resources.

Závěrem, realizace takového projektu vyžaduje skloubit **znalost zdravotnických standardů v ČR** (aby výstupy byly v souladu se všemi pravidly) s **moderními technologiemi** (API, bezpečnost, AI). Dobrou zprávou je, že mnoho potřebných stavebních kamenů už existuje: číselníky diagnóz a výkonů jsou dostupné, metodiky pojišťoven definují validační logiku a první prototypy AI pro kódování ukazují proveditelnost. Klíčové bude navrhnout systém tak, aby byl přínosem pro lékaře – tj. ulehčil jim administrativu, snížil chybovost vykazování a zároveň **nezdržoval jejich práci**, což obnáší důraz na ergonomii API a integrace (ideálně běžící automaticky na pozadí jejich stávajícího AIS). S důrazem na bezpečnost a správnost má toto řešení potenciál zaplnit důležitou mezeru v digitalizaci zdravotnictví – **automatizovat a zpřesnit komunikaci mezi poskytovateli a pojišťovnami**.