

# Sirius

Kompletní řešení pro práci s dokumenty a daty

## Sirius – kompletní řešení pro práci s dokumenty a daty

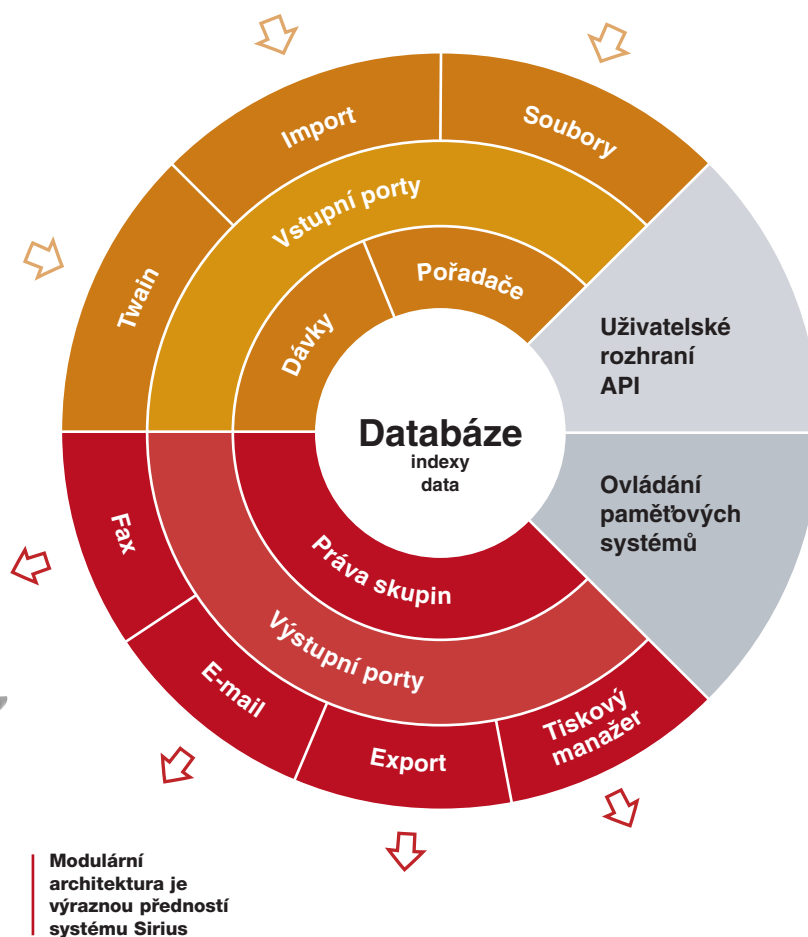
Sirius je modulární systém určený pro budování komplexních aplikací v oblastech, jako jsou digitalizace a správa dokumentů, systémy pro archivaci, zpracování a ukládání obrazových informací, optické rozpoznávání písma a zpracování formulářů. Používá se i v průmyslových aplikacích, které pracují s opticky získanými informacemi, například čtením čárových kódů apod.

Díky své databázově orientované architektuře klient/server a rozsáhlé sadě výkonných nástrojů dovoluje rychlou tvorbu robustních aplikací.



## Architektura

Celý systém je postaven v dvou a tříúrovňové architektuře klient/server nad standardní SQL databází. Klienti s grafickým uživatelským rozhraním pracují v prostředí Windows a webových prohlížečů.



## Modularita

Sirius je budován z jednotlivých funkčních prvků, které jsou obsaženy v programových modulech s jednotným rozhraním. Všechny funkce systému jsou důsledně řízeny propracovaným systémem přístupových práv, která jsou uplatňována pro konfiguraci systému, pro přístup uživatelů k jednotlivým funkcím a k uloženým informacím až na úroveň databázových polí. Je snadné vytvořit soubor funkcí na míru ušitý potřebám konkrétního uživatele. Standardní rozhraní dovoluje tvorbu nových funkčních modulů a rychlou realizaci zákaznických funkcí.

## Uživatelská práva

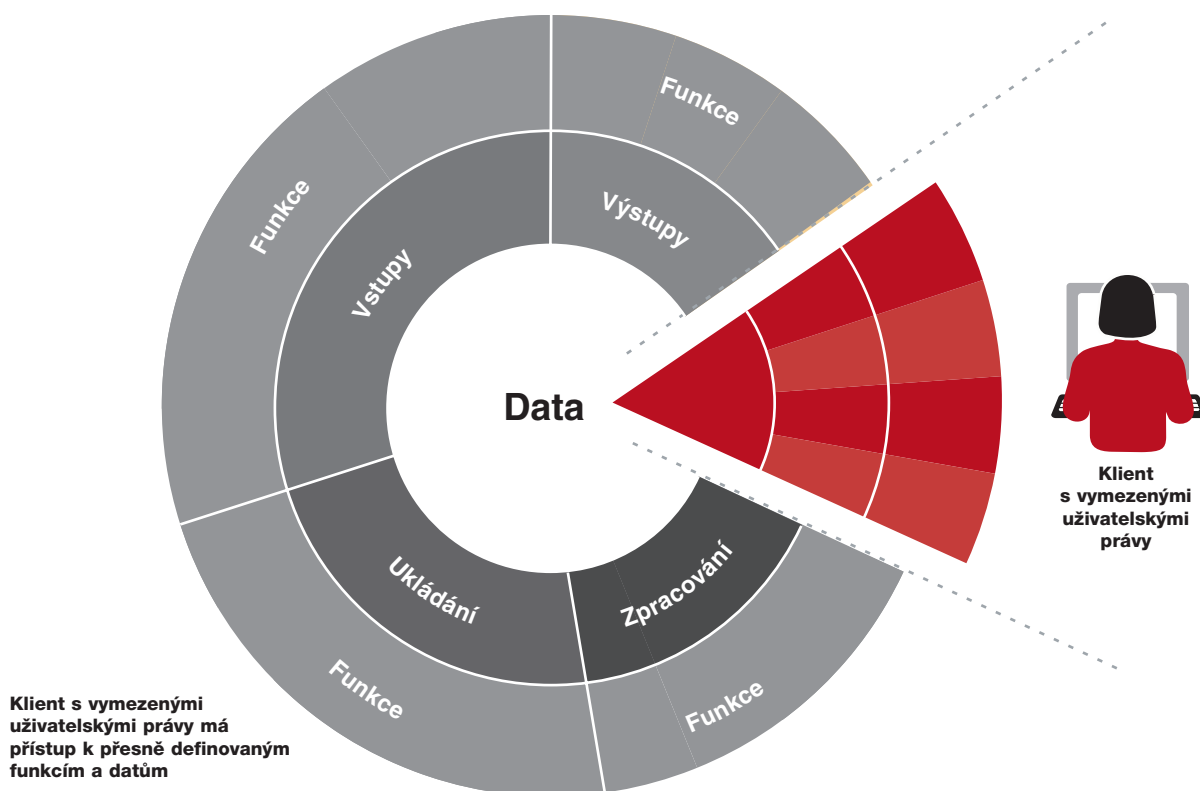
Uživatelská práva jsou definována jednak licenčními daty, která určují základní konfiguraci a funkčnost klienta a jednak právy, která jsou definovatelná pro jednotlivé uživatele nebo jejich skupiny. Uživatelská přístupová práva zpřístupňují nebo znepřístupňují jednotlivé elementární funkce klienta a jeho modulů a také definují přístupy k datům a objektům uloženým v databázi.

## Klient/server

### Klient

Modulární klient Siria pracuje v prostředí MS Windows s grafickým uživatelským rozhraním. Obsahuje rozhraní pro styk s prostředím a obsluhu v/v periferních zařízení. Integruje také databázového klienta a moduly třetích dodavatelů, jako např. OCR stroje apod. API rozhraní je určeno pro komunikaci s jinými aplikacemi, které nevyužívají technologii COM.

Tenký klient obsahuje všechny základní funkce tlustého klienta a umožňuje přístup přes web.



### Server

Klient komunikuje se standardním SQL databázovým strojem. Jsou podporovány databázové stroje MS SQL. Pro ukládání dat na archivační a zálohovací média jsou podporována standardní sw řešení obsluhující magnetopáskové, magnetooptické, CD a DVD jukeboxy.

## Funkční moduly

Stále rostoucí počet nových standardních i zákaznických funkčních modulů rozšiřuje funkcionalitu Siria. Jejich vhodnou volbou lze dosáhnout takové konfigurace, která přesně plní požadavky konkrétní aplikace.



### Vstupy

#### Základní skenovací modul

Jednodušší varianta skenovacího modulu je určena pro obsluhu skenerů se sw rozhraním TWAIN. Ovladače TWAIN dodává ke svým skenerům převážná většina světových výrobců.

#### Rozšířený skenovací modul

Rozšířený skenovací modul je vybaven pro dávkové zpracování velkého množství dokumentů na skenerech produkční kategorie a u podporovaných zařízení může obsluha ovládat veškeré funkce z klienta Siria. Spolupracuje plně s jeho dávkovým systémem. Uložené skenovací profily zrychlují práci při pořizování předem známých typů dokumentů.

#### Monitorování adresářů

Nové generace síťových skenerů a multifunkčních zařízení, jako jsou například digitální kopírky, dovolují jejich samostatné připojení k síti a přenos naskenovaných dokumentů prostřednictvím FTP na určené místo. Modul pro monitorování síťových adresářů dovoluje systému Sirius automaticky sbírat naskenované dokumenty a zpracovávat je stejnými metodami jako dokumenty z ostatních zdrojů.

#### Import z jiných aplikací

Přenos dokumentů z jiných aplikací je možný programově prostřednictvím API, metodou drag and drop z průzkumníka Windows nebo pomocí modulu import. Importované objekty zůstávají ve svém původním formátu. To umožňuje vytvářet kombinované dokumenty, tj. ukládat a indexovat dokumenty digitalizované z papírové podoby spolu s dokumenty, které vznikly v elektronické podobě – např. výstupy kancelářských a obchodních aplikací.

data	celkem	...	...
16.03.02	14	...	...
17.03.02	12	...	...
celkem	26	...	...
05.04.02	17	...	...
16.04.02	9	...	...
17.04.02	9	...	...
18.04.02	9	...	...
19.04.02	12	...	...
22.04.02	3	...	...
23.04.02	12	...	...
24.04.02	12	...	...
29.04.02	12	...	...
celkem	90	...	...
10.05.02		...	...
...	...	...	...



## Zpracování

### Dávkové zpracování

Propracovaný systém dávkového zpracování je určen pro distribuované zpracování středních a velkých množství dokumentů. Dokumenty jsou po celou dobu trvání procesů zpracování uloženy v tzv. dávkách od svého pořízení prostřednictvím některého ze vstupních portů až po uvolnění zpracovaných dokumentů do systému, kterým dávkové zpracování končí. Jednotlivé dílčí úlohy nad dokumenty v dávce mohou být prováděny na specializovaných pracovištích, např.: skenování, rozpoznávání a čtení dokumentů nástroji OCR/ICR, verifikace vytěžených dat a indexování pro archivaci.

### Zpracování obrazu

Sirius podporuje ve všech svých modulech práci s dokumenty v jejich původním přirozeném formátu. Zvláštní pozornost je věnována grafickým formátům u dokumentů pořizovaných skenováním. Jsou používány specifické standardizované formáty vhodné pro digitální archivaci dokumentů především v černobílé podobě, které jsou optimalizovány pro optimální poměr mezi kvalitou zobrazení a velikostí dokumentů. Jsou používány bezztrátové kompresní algoritmy a inteligentní postupy pro zobrazování dokumentů na počítačových displejích (scale to gray).

Zpracování obrazu umožňuje podstatné zlepšení čitelnosti skenovaných dokumentů pro rozpoznávací stroje OCR/ICR a významně snižuje chybovost při vytěžování dat.

Funkce zpracování zahrnují především operace, které zajistí korektní registraci dokumentu, tj. nalezení okrajů a vyrovnání dokumentů, odstranění nečistot a prvků, které nenesou textovou informaci, jako jsou rámečky, rastry apod.

### Rozpoznávání formulářů

Sirius umožňuje rozpoznávat předem definované typy dokumentů (formulářů) při jejich vstupu do systému nebo v dalších krocích jejich zpracování. Dokumenty jsou rozpoznávány podle specifických znaků, které jsou jedinečné pro daný typ dokumentu. Každému typu dokumentu je možno přiřadit tzv. šablonu, která definuje, jakým způsobem se s daným typem nakládá v průběhu zpracování. Šablona rozděluje dokument na zóny pro čtení nástroji OCR/ICR. Pro každou zónu určuje kromě jejího umístění na formuláři a rozměrů také typ informace, který tato zóna obsahuje, přiřazuje databázové pole, kam se ukládají vytěžená data a definuje kontrolní mechanismus pro verifikaci těchto dat.

**Pořizování dat s podporou  
Siria snižuje pracnost  
až o 50 procent**



Rozpoznání typu dokumentu dovoluje zpracovávat bezchybně „směsí“ dokumentů, kde jsou v jedné dávce pomíchány různé druhy dokumentů, které mají buď charakter formulářů a jsou určeny především pro vytěžování dat, nebo jde o dokumenty, jejichž formát není předem znám a s jejichž obsahem se zachází ad hoc.

### **OCR/ICR/OMR/čárové kódy**

Nástroje Siria pro vytěžování dat jsou založeny na výkonném rozpoznávacím stroji pro optické čtení skenovaných dokumentů. Jsou rozpoznávány všechny běžné typy strojového písma (OCR), ručně psané hůlkové písmo (ICR), zaškrťovací okénka a čárové kódy. Jsou podporovány všechny východoevropské jazyky a znaky s diakritikou. Čtení dokumentů je možné buďto podle šablon rozeznávaného typu dokumentu, nebo celostránkově s vazbou na fulltextovou databázi. Specifické funkce pro čárový kód umožňují např. jeho nalezení na stránce bez ohledu na to, kde je umístěn, rozeznání jeho typu apod.

Dlouholeté zkušenosti firmy Elsys Engineering se zpracováním obrazu umožňují realizovat specializované zákaznické úlohy pro vytěžování dat, založené na analýze obrazové informace z různých zdrojů.

### **Vytěžování dat/indexování dokumentů**

Jednou z hlavních úloh při zpracování dokumentů je indexování, to znamená připojit k dokumentu data, která jej popisují a slouží pro uložení a především pro vyhledávání dokumentů. Vytěžování dat z dokumentů nahrazuje tradiční způsob pořizování dat, která slouží jako vstupní data pro jiné informační systémy. Indexování a vytěžování lze v prostředí Siria provádět v jednom pracovním chodu. Data vytěžená z dokumentů jsou použita jako zdroj pro informační

systemy a současně jsou ukládána do indexových databází Siria. Sirius poskytuje velmi efektivní podporu pro pořizování dat z naskenovaných dokumentů i v takových případech, kdy se jedná o informace strojově nerozpoznatelné a podstatně omezují manipulaci s papírovými dokumenty (Key from image). Tato metoda snižuje pracnost při pořizování dat až o 50 %.

### **Indexy**

Indexy jsou záznamy v databázi. Každý dokument může mít libovolný počet různých indexů. Jejich struktura umožňuje přehledné hierarchické strukturování dokumentových databází. Jejich obsahy jsou plněny nástroji pro vytěžování dat z dokumentů nebo ručně. Přístup k indexům a k jejich editování je stejně jako všechny funkce Siria řízen přístupovými právy.





## Ukládání dokumentů v databázi

### Organizace databáze

Databáze Siria je vybudována na standardní relační SQL databázové platformě. Standardně jsou podporovány běžné databázové stroje, jako je MS SQL apod. V databázových tabulkách jsou ukládány všechny informace, které se vztahují ke zpracovávaným dokumentům, ale také samotné dokumenty a všechna procesní data Siria a jeho klientů. Klient Siria pracuje s jednou nebo více různými databázemi.

### Úložiště dokumentů

V případech, kdy je to účelné, je možno ukládat dokumenty do externích úložišť nebo pracovat s dokumenty uloženými distribuovaně na místě jejich vzniku. Je podporováno uložení v souborovém systému OS a rozhraní pro specializované souborové systémy pro řízení velkokapacitních hierarchických paměťových struktur, jako jsou disková pole, jukeboxy a offline úložiště a zálohovací nástroje a jejich kombinace.

### Struktura indexů

Strukturu indexů a jejich počet volí uživatel Siria podle aplikace. Indexy jsou plně editovatelné a je možno měnit jejich strukturu a doplňovat nové indexy i dodatečně, bez nároků na znalost databázových nástrojů.

### Vyhledávací nástroje

Vyhledávání je při práci s dokumenty nejčastější činnost, která také v klasické papírové podobě zabírá nejvíce času. Proto jsou v Siriu integrovány mimořádně výkonné vyhledávací nástroje, které uspokojí všechny kategorie uživatelů. Je samozřejmě možné vyhledávat podle všech indexových polí dokumentu nebo podle obecných vlastností hledaného objektu, jako jsou velikost, formát, barevnost apod. Pro konstrukci složitějších vyhledávacích dotazů je v klientu Siria k dispozici pomocník, který s uživatelem navrhne i velmi složitý dotaz a uloží jej pro opakované použití. Uložené

dotazy je možné vždy znovu použít, případně upravit při změně zadání. Výběr dokumentů, který je výsledkem vyhledávání, je možno použít pro další např. dávkové zpracování.

### Zobrazovací nástroje

Součástí klienta je multiformátový prohlížeč, který dovoluje zobrazovat všechny běžné dokumentové a grafické formáty. Pro ostatní typy dokumentů je možno připojit specializované prohlížeče.

### Podpora archivačních medií

Pro aplikace, které mají charakter digitální archivace a požadují splnění specifických požadavků na tzv. pravou archivaci, to znamená např. dlouhodobou trvanlivost a nemodifikovatelnost zpracovávaných a uložených dokumentů, poskytuje Sirius podporu hw s potřebnými vlastnostmi, jako jsou např. magnetooptická média typu WORM a poskytuje procesní ochranu, která brání modifikaci dokumentů nebo neoprávněnou manipulaci s nimi v průběhu jejich zpracování.

### Zálohování

Zálohování je samozřejmou součástí každé bezpečnostní strategie při práci s dokumenty. Sirius využívá pro zálohování zabudované zálohovací mechanismy, které umožňují plánované zálohování na zálohovací média. V instalacích s hierarchickými úložišti jsou využívány mechanismy, které zajišťují automatické odkládání dokumentů určených pro archivaci na archivační média a jejich zálohování.

### Podporovaný storage hw/sw

Sirius spolupracuje se standardními produkty světových výrobců pro ukládání dat, jako je např. HP, Tandberg, Plasmon a s doprovodnými softwarovými produkty specializovaných výrobců.



## Výstupy

### Tiskový manažer

Tiskový manažer slouží k formátování tiskových výstupů z dokumentových databází, tisku personifikovaných dokumentů a formulářů, které kombinují databázová data s určitým formulářem.

### Export

Exportní funkce Siria slouží pro předávání vytěžených dat, databázových informací Siria a dokumentů uložených v Siriu jiným aplikacím v široké paletě formátů. Speciální funkcí je export části databáze Siria na přenosná média (např. CD) včetně prohlížečícího klienta. Tato funkce dovoluje distribuovat části dokumentových databází i uživatelům, kteří nejsou vybaveni Siriem.

### Distribuční nástroje: mail, fax, tisk

Klient Siria obsahuje standardně nástroje pro distribuci dokumentů, tj. jejich tisk, faxování přímo z prostředí klienta a odesílání vybraných dokumentů elektronickou poštou.

### Podporovaná výstupní zařízení

Sirius podporuje standardní zařízení operačního systému. S výhodou je možno využít zařízení s podporou přímého tisku grafických formátů, jako jsou např. tiskárny s podporou TIFF a PDF v aplikacích s velkokapacitními tisky.



## Integrace Siria s jinými aplikacemi

### Podporovaná rozhraní a standardy

Sirius podporuje standardní souborové formáty, které zajišťují bezproblémovou výměnu dat s jinými standardními aplikacemi.

Spolupráce a integrace Siria s jinými aplikačními systémy je snadná díky objektové architektuře a důslednému uplatnění technologie COM. Integrátoři mohou využít také knihovnu API, která poskytuje všechny funkce pro přístup do databází Siria.

V prostředí průzkumíka Windows je možné vkládat soubory do Siria metodou drag and drop.

Soubory ze standardních kancelářských aplikací jako je např. MS Word nebo MS Excel je možno ukládat do Siria přímo z těchto aplikací.

### Prohlížeče dokumentů

Zobrazování dokumentů v integrovaných aplikacích je snadné díky autonomnímu prohlížeči, který zobrazuje všechny formáty dokumentů Siria.

### Ochrana dat

Pro přenos dat mezi Siriem a jinými systémy a pro ochranu dat uložených na off-line médiích a zálohách je pro kritické aplikace možno použít funkce pro šifrování dat, které ochrání důležitá data proti zneužití.

## Cílové aplikace Siria a jejich hlavní výhody

### Aplikace pro skenování a zpracování obrazu

Pomocí systému Sirius lze budovat aplikace pro digitalizaci dokumentů ve spolupráci se skenery různých tříd a výrobců podle volby uživatele a jejich efektivní předzpracování do formátů optimálních pro další použití pro digitální archivaci, distribuci a vytěžování dat.



*Optimální spolupráce hardwaru a softwaru díky dokonalému přizpůsobení Siria hardwarovým rozhraním.*

### Integrace správy dokumentů do informačních systémů

Díky širokému spektru možností a rozhraní pro předávání informací, je možno začleňovat systém Sirius do už existujících informačních systémů a poskytovat jejich uživatelům specifické služby, jako je například vyhledávání dokumentů v digitálním archivu a jejich zobrazení, vstup (skenování) papírových dokumentů a jejich ukládání spolu s tradičními strukturovanými informacemi informačního systému.



*Stejný přístup ke všem typům informací, jak digitálním tak „původně papírovým“.*

### Oběh dokumentů

Řízení toku dokumentů v aplikacích typu spisová služba, došlá pošta apod. s přímým přístupem účastníků k digitalizovaným dokumentům, sledování plnění úkolů, upozorňování a podobně.



*Odstranění fyzické manipulace s dokumenty z pracovních procesů.*

### Dokument management

Kompletní péče o dokumenty v jejich digitální podobě bez ohledu na formát nebo způsob pořízení během celého jejich životního cyklu, podpora jejich tvorby, využívání a aplikace.



*Kladné účinky toho, že jsou všechny informace pod kontrolou se projeví především v systémech řízení kvality.*

## Archivace dokumentů

---

Technologie pro ukládání digitalizovaných dokumentů, která splňuje všechny požadavky na „pravou archivaci“ tj. dlouhodobost, bezpečnost a nemodifikovatelnost bez nechtěných změn klasické archivace, jako jsou velké nároky na prostor, nákladné ukládání i vyhledávání apod.



*Mimořádné úspory při práci s archivem.*

## Zpracování formulářů, sběr a pořizování dat

---

Zcela nebo částečně automatizované získávání dat z dokumentů, které mají charakter buďto formulářů nebo volného textu a jejich efektivní verifikace a předávání do cílových databází. Podpora ručního a strojového písma, různých typů značek a všech běžných typů čárových kódů.



*Podstatné zvýšení produktivity při pořizování dat z papírových dokumentů.*

## Obrazová databáze

---

Specializované funkce pro zpracování barevných digitálních fotografií, tvorba alb, indexování pomocí klíčových slov, distribuce na CD.



Elsyst Engineering je společnost zaměřená především na vývoj vlastních softwarových produktů a realizaci kompletních projektů a služeb v oboru správy a zpracování dokumentů, průmyslové automatizace a zakázkové mikroelektroniky. Rozsáhlé jsou její aktivity v oblasti softwaru pro digitální fotografii a v produkci grafického designu.

**Elsyst  Engineering**

Elsyst Engineering » Brněnská 10, 682 01 Vyškov  
tel.: +420 517 334 070-2 » fax: +420 517 334 071  
ee@ee.cz » www.ee.cz