



Planet Pharma Solutions

PHARMADOC USERS CONFERENCE 2023

文書間リンク作成ツール PharmaDoc JetLinkのご紹介

2023年06月02日

プロジェクトアドバイザー 都丸 淳之様

プラネットファーマソリューションズ株式会社 大平 聡、古谷 寿佳

国際大学GLOCOM客員研究員 小野塚 亮様

アジェンダ

- PharmaDoc JetLink 開発の背景と目的
- 文書間リンクの自動作成の試作の結果報告
- 開発中の文書間リンク作成ソリューション
「PharmaDoc JetLink」のご紹介
 - 製品の特長
 - 開発中イメージ
 - 他のシステムとの連携

PharmaDoc JetLink 開発の背景と目的

文書間リンクに関する要件 (ICH SSF, JP SSF)

[別紙4 \(ICH 電子化コモン・テクニカル・ドキュメント \(eCTD\) に含める電子ファイル仕様 v1.3\)](#)

2.13 文書のナビゲーション (ハイパーテキスト・リンク、ブックマークおよび目次)

- ハイパーテキスト・リンクは、PDF文書のナビゲーションを向上させるテクニックである。
- 文書全体を通して、同一ページ上にない関連セクション、参考文献、付録、表または図へのハイパーテキスト・リンクを設定することは有用であり、ナビゲーションの効率を向上させる。

[別紙2 \(ICH 電子化コモン・テクニカル・ドキュメント \(eCTD\) に含める電子ファイル仕様の国内実装について v1.3.0\)](#)

4.3 ハイパーテキスト・リンク

- 原則としてICH SSFの記載に基づくこと。同一文書内 (同一PDF ファイル) 及び異なる文書間 (異なるPDFファイル) のハイパーテキスト・リンクは、審査の効率化のために適切に設定すべきものである。第2部から第3部～第5部へのハイパーテキスト・リンクは可能な限り設定すること。

SSF: Specification for Submission Formats for eCTD

文書間リンク作成の高速化に着目した理由

文書間リンク作成には長い時間が掛かる！

「作成・自己点検+第三者点検」で1本あたり3分だと、1000本で50時間。

本数	時間	もしも25分の1にできたら
1000本	50時間	2時間
2000本	100時間	4時間
...
5000本	250時間	10時間

Module 5.3.5.1, PPS-005 CSR, Section 12.2.1

Module 5.3.5.2, PPS-008 CSR, Appendix 16.2.8, Listing 8.6

ひとつのシナリオ；最初のマイルストーン



	Regulatory Operations 2年目	SME (特定分野の専門家)
JetLink	○	—

Module 5.3.5.1, PPS-005, CSR, Section 12.2.1
 Module 5.3.5.2, PPS-008, CSR, Appendix 16.2.8, Listing 8.6

[eCTD v4.0] codeSystem Name: **ICH Context of Use**

Code	Description	Keyword Type that can be applied to this CoU
ich_5.3.5.2	m5.3.5.2 study reports of uncontrolled clinical studies	study id_study title (R), document type (R), indication (R), site-id (O), study group order (O), group title (O)

文書間リンクの自動作成の試作の結果報告

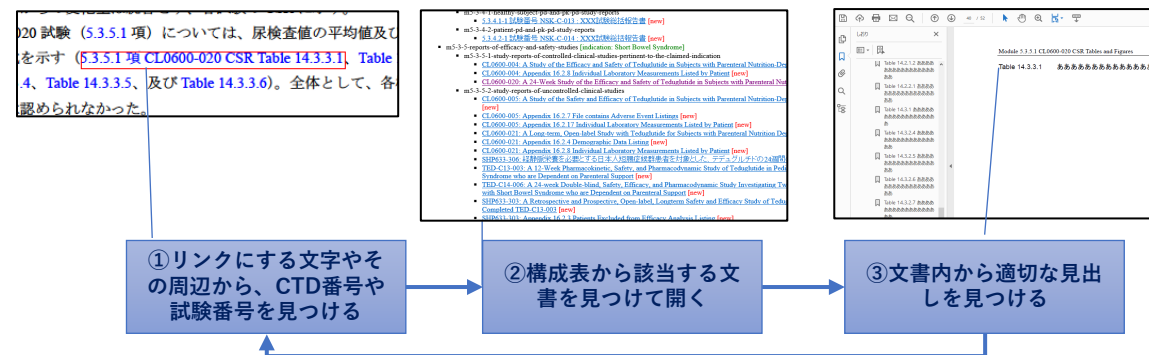
はじめに

- 昨年はeCTDの情報を利用して、PDFリーフファイルの文書間リンクを自動検索するシステムのデモを見ていただきました。
- その後、製薬企業各社にご協力をいただき、実物のeCTDを使用した自動作成システムの精度検証を行いました。
- ご協力いただいている製薬企業各社および担当者の皆様には、この場を借りて厚く御礼申し上げます。



文書間リンクの自動作成について

- 文書間リンクの作成作業は、SR化作業の中でも、依然負担の大きな作業として残っています。
- 弊社は自然言語処理AIを利用することで、文書間リンクの飛び先の探索を高精度に実現できると考え、昨年から技術検討を行ってきました。



飛び先の検索をAIで行う

- 文章の中から特定の属性の文字列を抽出する
 - CTD番号
 - 試験番号（資料名）
 - 図表番号（項番号）

→文書処理技術「**固有表現抽出**」を利用できる

※固有表現抽出とはテキスト中から人名、地名、組織名、日付、時間、数量など、特定の固有物を自動的に抽出する技術のこと



デモアプリでの飛び先検索実験（昨年）

クライアントアプリ

CL0600-020 試験 (5.3.5.1 項) 及び CL0600-004 試験 (5.3.5.1 項) については、尿検査値及びベースラインからの変化量は統合せず、各試験の CSR に示す。
 CL0600-020 試験 (5.3.5.1 項) については、尿検査値の平均値及び中央値のベースラインからの経時変化を示す (5.3.5.1 項 CL0600-020 CSR Table 14.3.3.1, Table 14.3.3.2, Table 14.3.3.3, Table 14.3.3.4, Table 14.3.3.5, 及び Table 14.3.3.6)。全体として、各検査項目に臨床的に意味のある群間差は認められなかった。
 CL0600-004 試験 (5.3.5.1 項) の尿検査値の経時変化を 5.3.5.1 項 CL0600-004 CSR Section 12.4.1.3 及び 5.3.5.1 項 CL0600-004 CSR Table 14.3.4.2.84, Table 14.3.4.2.85, Table 14.3.4.2.86,

①テキストを抽出

CL0600-020 試験 (5.3.5.1 項) については、尿検査値の平均値及び中央値のベースラインからの経時変化を示す (5.3.5.1 項 CL0600-020 CSR Table 14.3.3.1, Table 14.3.3.2, Table 14.3.3.3, Table 14.3.3.4, Table 14.3.3.5, 及び Table 14.3.3.6)。全体として、各検査項目に臨床的に意味のある群間差は認められなかった。

ファイル名: .../cl0600-020-report-body.pdf
 図表番号: Table14.3.3.1

④図表項番号の飛び先座標を検索

ファイル名: .../cl0600-020-report-body.pdf
 ページ番号: 40
 座標: (70, 769)

AIサーバー

CL0600-020 試験 (5.3.5.1 項) については、尿検査値の平均値及び中央値のベースラインからの経時変化を示す (5.3.5.1 項 CL0600-020 CSR Table 14.3.3.1, Table 14.3.3.2, Table 14.3.3.3, Table 14.3.3.4, Table 14.3.3.5, 及び Table 14.3.3.6)。全体として、各検査項目に臨床的に意味のある群間差は認められなかった。

②AIで固有表現を抽出 (CTD番号、試験番号、図表項番号)

CTD番号: 5.3.5.1 項
 試験番号: CL0600-020
 図表番号: Table14.3.3.1

③構成表から飛び先ファイルを特定

CTD番号	試験番号	ファイル名
5.3.5.1 項	CL0600-020	.../cl0600-020-report-body.pdf
...

ファイル名: .../cl0600-020-report-body.pdf
 図表番号: Table14.3.3.1

学習データは別途準備

デモアプリの性能評価（昨年発表）

■精度

192件のリンクに対し**90%以上の精度**で正しい飛び先を作成することができた
※擬似的なeCTDによる参考値

■処理速度

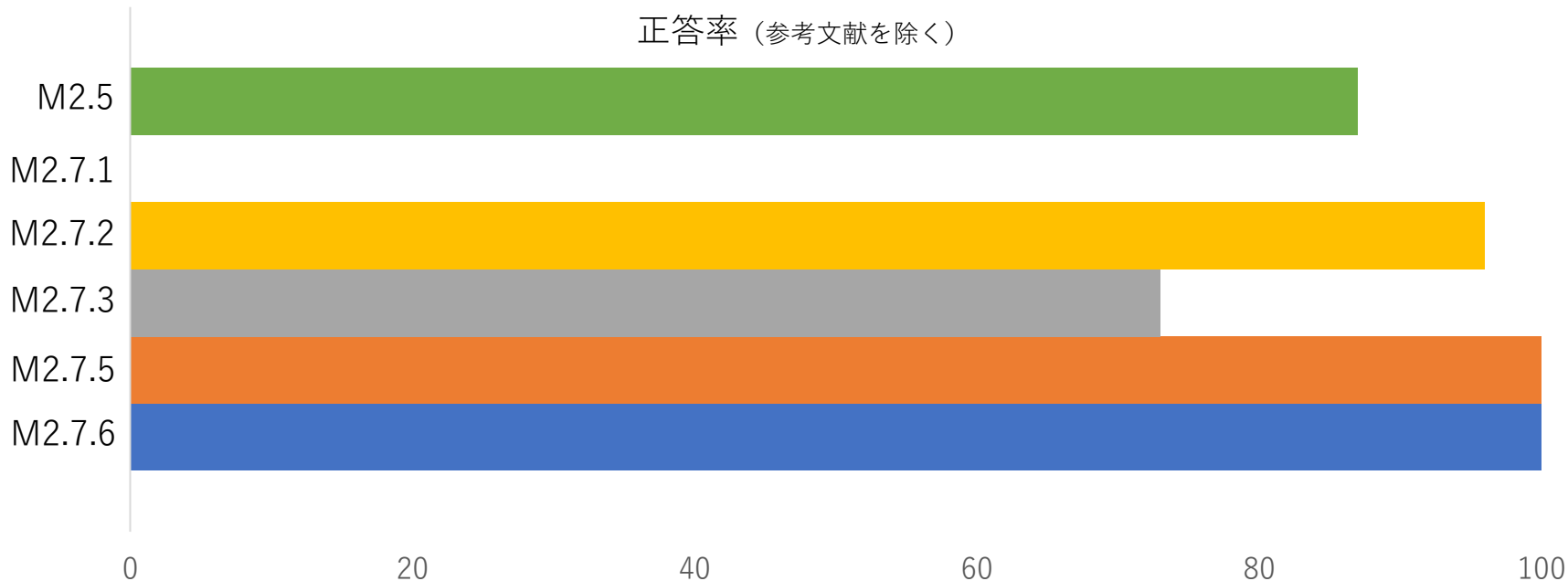
リンクの飛び先抽出が約11秒、PDFへの反映に約17秒
手作業に比べ**圧倒的に高速**に処理することが可能



実際の申請文書を用いた性能評価（現時点）

■精度

文書の記述方法によりばらつきがみられるが、固有表現抽出によるAI処理は安定して高い性能を示した。A社ご提供のeCTDでは、M2の文書に対して**88%の正答率**でリンクの設定が行えた（下図参照）。



■速度

PDFファイルのサイズによって処理時間に差はでるものの、手作業に比較しての優位性は維持した。

デモアプリの外部評価により示された課題

- AIの教師データ作成が煩雑で、かつ学習の効果が見えにくい
⇒**最小限の教師データで効果を発揮する探索手法への変更**
- 抽出したCTD番号や図表項番号だけでは正しい飛び先に到達できないケースがある
 - 参考文献
 - リンク元と飛び先での表記の違い
(例：リンク元「表 1」、飛び先「Table1」)
 - 飛び先ファイルが複数に分割されている など⇒**探索ロジックの改良**
- 文書間リンクの手動作成・編集機能を持たない
 - 100%でない以上、自動作成したリンクを未確認で採用は難しい
 - 飛び先が正しいことを確認・修正する作業のほうが負担が大きそう⇒**編集のためのUIの充実**



JetLinkの開発計画

- 第一期
 - 文書間リンクの飛び先検索機能の技術検証（評価機）
- 第二期
 - デスクトップアプリの評価版提供 **2023年10月を予定**
 - デスクトップアプリ製品リリース
 - 疑似リプレース機能追加
- 第三期
 - Web版製品リリース



開発中の文書間リンク作成ソリューション 「PharmaDoc JetLink」のご紹介

PharmaDoc JetLink

- 目的

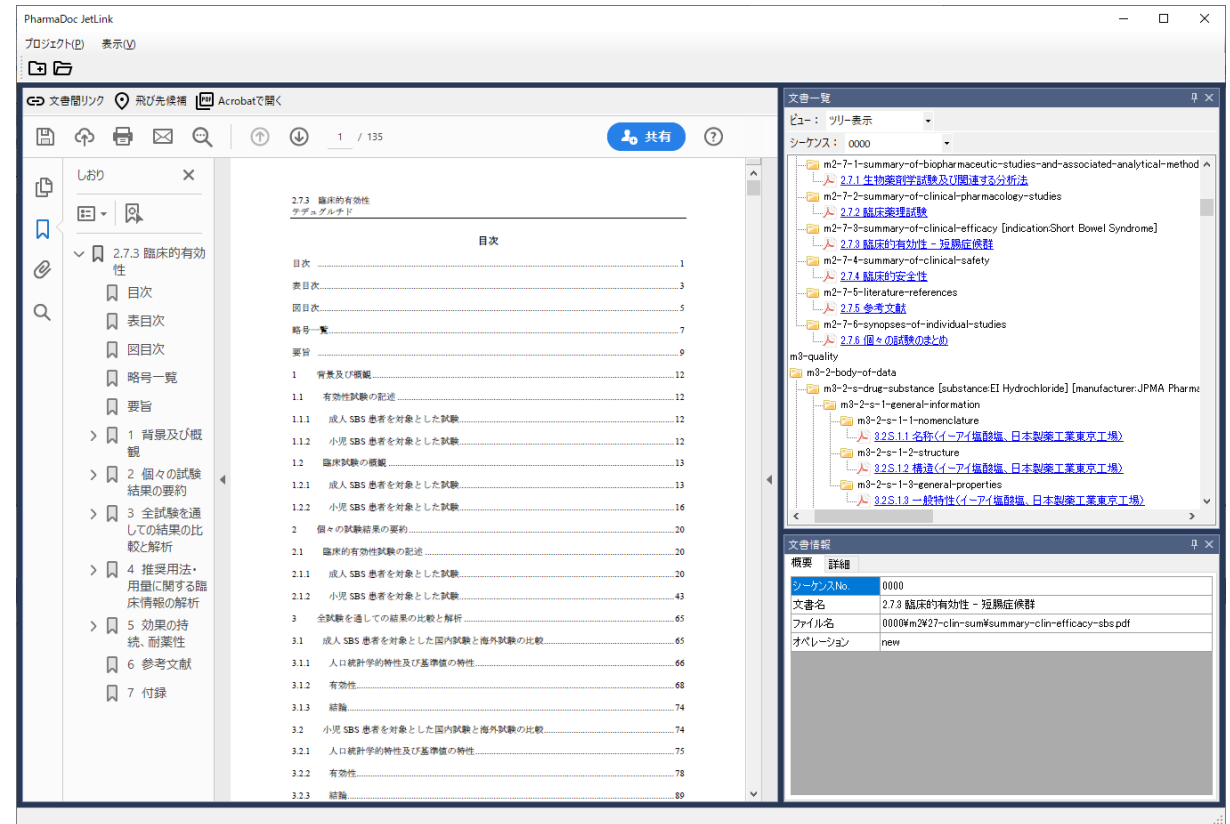
- eCTD編纂作業や申請ライフサイクルに関わる作業におけるリーフファイルの文書間リンク作成を効率的に行う

- 機能

- 文書間リンクの作成・編集機能 (Ver.1.0)
 - リンク枠の一括作成機能
 - 飛び先の一括付与機能
- ライフサイクル対応 (Ver.2～)
 - 疑似リプレース

- 動作環境

- Windows 10/11
- Acrobatは不要



PharmaDoc JetLinkの三つの特長

① AIの補助による飛び先の自動付与

AIを利用して、CTD構造に基づいた飛び先を自動検索することにより、文書間リンク作成作業のストレスを大幅に軽減します。

② eCTD構造に基づいたわかりやすいUI

CTD構造に基づいた文書のツリーとXMLメッセージ内の詳細情報の表示により、eCTD内のリーフファイルの役割を意識したリンク作成作業が可能です。

③ JetLink DBによるリンク情報の一元管理

ローカルデータベースを使ってeCTDパッケージ内のリンクデータを一元管理することで、複雑化しがちな文書間のリンク管理を扱いやすくします。



特長① AIの補助による飛び先の自動付与

飛び先検索AIの搭載により、リンク枠の適切な飛び先を検索します。

- 編集画面でユーザーがリンク枠の飛び先を選択する際に、適切な飛び先をAIがサジェストします。
- 複数のリンク枠に対して、飛び先をまとめて付与することができます（一括付与）。



(参考) JetLinkにおける検索ロジックの改良

1. 固有表現抽出手法の変更

新手法「フレーズマッチャー」の採用により、以下を実現しました。

- ローカル環境での動作
- 処理の高速化
- 教師データ作成の作業軽減

2. 多段階検索

飛び先文書内に適切な見出しが見つからない場合、さらにそこからリンクするファイルを探します。

3. JetLink構成表

eCTDのXMLメッセージに記載されない情報を補足することで、より正確な飛び先文書検索を可能にします。

- リーフファイルのグループ化 (分割したファイルの特定)
- ドキュメントタイプ (Ver.3.2.2の場合)
- 文書名の別名

ライブラリ名	フォルダ階層	XMLエレメント名	小タイトル	ファイル名	サイズ(MB)	MD5	検索パス	検索
		md-R001						
		md-R002						
		md-R003						
		md-R004						
		md-R005						
		md-R006						
		md-R007						
		md-R008						
		md-R009						
		md-R010						
		md-R011						
		md-R012						
		md-R013						
		md-R014						
		md-R015						
		md-R016						
		md-R017						
		md-R018						
		md-R019						
		md-R020						
		md-R021						
		md-R022						
		md-R023						
		md-R024						
		md-R025						
		md-R026						
		md-R027						
		md-R028						
		md-R029						
		md-R030						
		md-R031						
		md-R032						
		md-R033						
		md-R034						
		md-R035						
		md-R036						
		md-R037						
		md-R038						
		md-R039						
		md-R040						
		md-R041						
		md-R042						
		md-R043						
		md-R044						
		md-R045						
		md-R046						
		md-R047						
		md-R048						
		md-R049						
		md-R050						
		md-R051						
		md-R052						
		md-R053						
		md-R054						
		md-R055						
		md-R056						
		md-R057						
		md-R058						
		md-R059						
		md-R060						
		md-R061						
		md-R062						
		md-R063						
		md-R064						
		md-R065						
		md-R066						
		md-R067						
		md-R068						
		md-R069						
		md-R070						
		md-R071						
		md-R072						
		md-R073						
		md-R074						
		md-R075						
		md-R076						
		md-R077						
		md-R078						
		md-R079						
		md-R080						
		md-R081						
		md-R082						
		md-R083						
		md-R084						
		md-R085						
		md-R086						
		md-R087						
		md-R088						
		md-R089						
		md-R090						
		md-R091						
		md-R092						
		md-R093						
		md-R094						
		md-R095						
		md-R096						
		md-R097						
		md-R098						
		md-R099						
		md-R100						

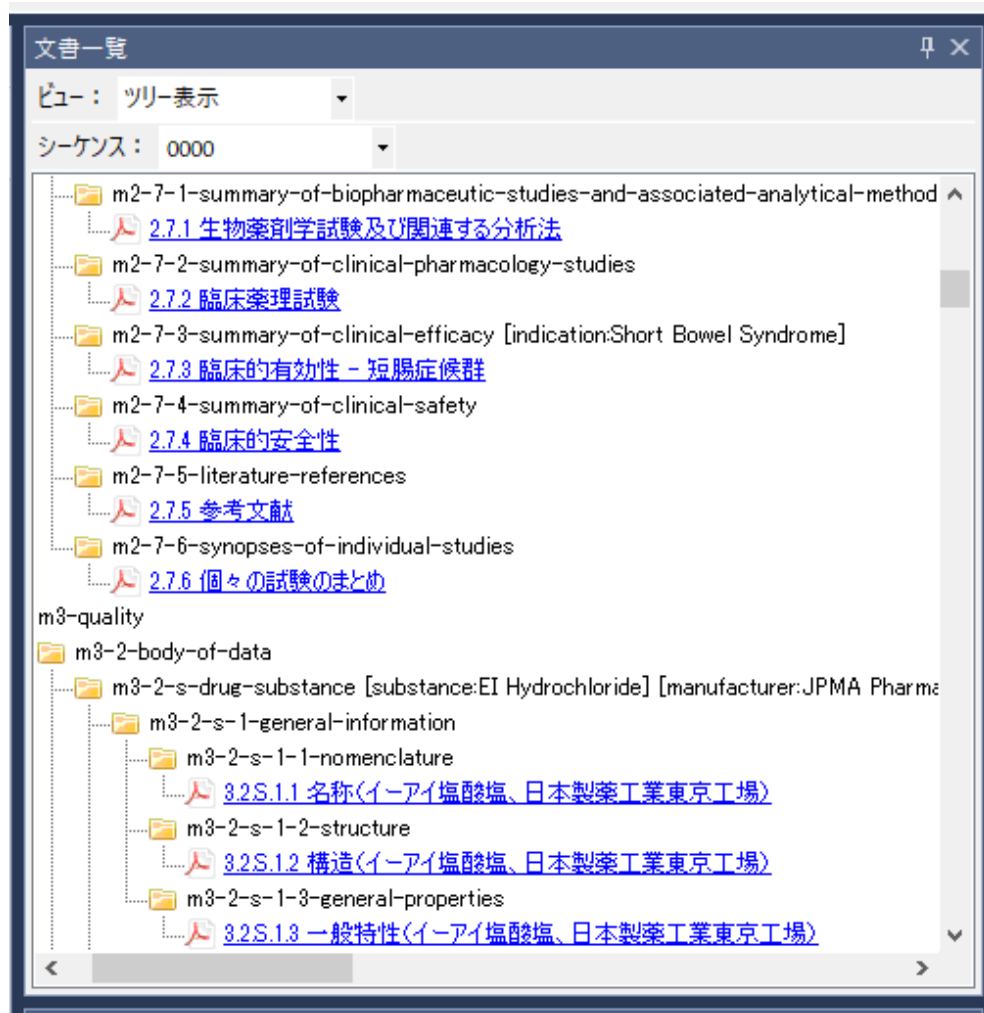
(参考)AI技術の概要

- 自然言語処理の一分野である固有表現抽出を行う
- 固有表現抽出とはテキスト中から人名、地名、組織名、日付、時間、数量など、特定の固有物を自動的に抽出する技術のこと
- 本アプリケーションではフレーズマッチャーという技術を利用している



- フレーズマッチャーでは元のテキストから自動的にパターンを抽出して、新しいテキストに対してパターンマッチングを行うことができる
- 教師データの準備が容易である点、実行速度が速い点に利点がある

特長② eCTD構造に基づいたわかりやすいUI



- 簡易的なeCTDビューア機能を搭載し、CTDの構造やリーフファイルのメタデータを確認しながら作業することができます。
- eCTD v4.0にも対応します。

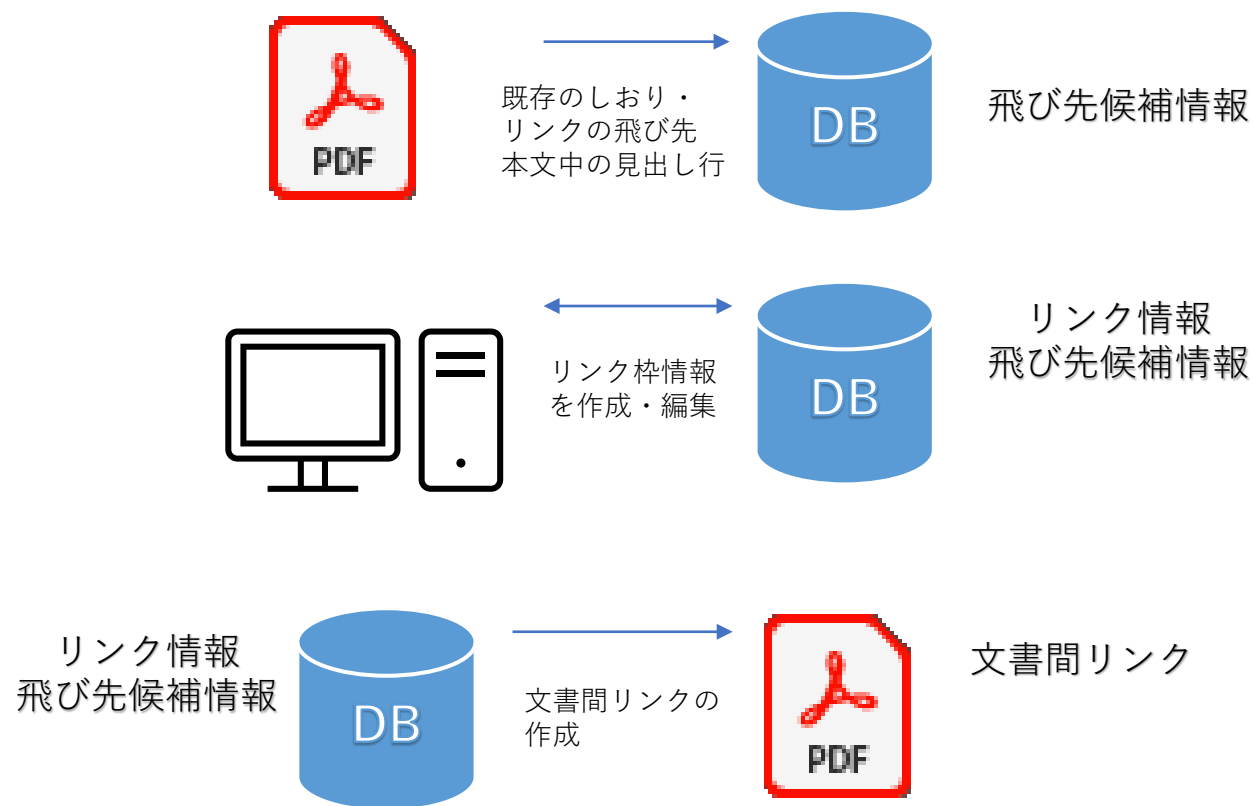


特長③ JetLink DBによるリンク情報の一元管理

ローカルのデータベースを用いて文書間リンクの作成作業を仮想化します。

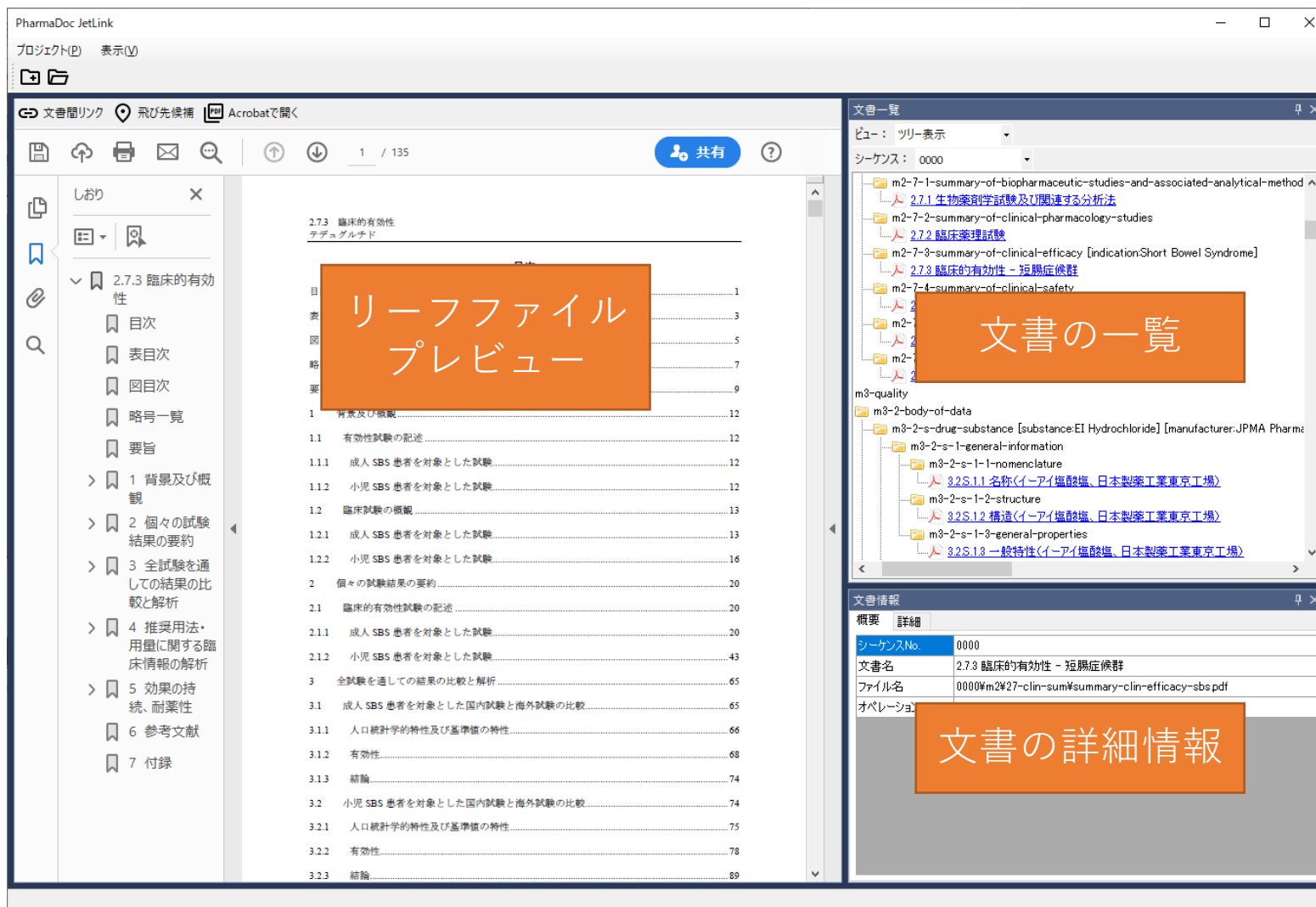
リンク情報を一元管理することで、様々な変更作業を統合的に行うことができます。

1. eCTDパッケージごとに、ローカルデータベース（JetLink DB）を作成します。
2. ユーザーが作成・編集したリンク枠や飛び先の情報は、直接PDFに反映せず、データベース内に蓄積されます。
3. ユーザーの明示的な操作により、PDF上に実際にリンクが作成されます。



JetLinkの開発中イメージ

1. メイン画面



2. 文書間リンクの編集画面

文書間リンクの編集
新規 削除 リンク枠の一括作成 飛び先の一括付与 PDFに反映

リンク枠の一覧

No.	ページ	枠内文字列
43	2	7 付録
44	13	5.3.5.2項(final)SHP633-301 final CSR...
45	21	Module 5.3.5.1, CL0600-020 CSR, Figu...
46	22	Module 5.3.5.1, CL0600-020 CSR Tab...
47	22	Module 5.3.5.1, CL0600-020 CSR Fig...
48	24	Module 5.3.5.1, CL0600-004 CSR, Figu...
49	25	Module 5.3.5.1 CL0600-004 CSR Tabl...
50	25	Module 5.3.5.1 CL0600-004 CSR, Tab...
51	26	Module 5.3.5.1 CL0600-004 CSR, Tabl...
52	26	5.3.5.1項CL0600-004 CSR
53	26	Table 14.2.1.1.
54	27	5.3.5.1項CL0600-004 CSR Section 11...
55		Fig...
56		Fig...
57		Fig...
58		Fi...
59		Fig...
60	44	Module 5.3.5.2 TED-C13-003 CSR, Fi...
61	46	Module 5.3.5.2 TED-C14-006 CSR, Fi...
62	48	Module 5.3.5.2 SHP633-301 CSR, Fig...
63	50	5.3.5.2項(final)
64	50	SHP633-301 final CSR Section 10.1
65	52	Module 5.3.5.2 SHP633-303 CSR, Fig...
66	53	Module 5.3.5.2 SHP6343-303 CSR, Fig...
67	55	Listing 16.2.4R
68	55	Listing 16.2.3.1R
69	55	SHP633-303 CSR, Table 14.2.2.1R

リンク枠プレビュー

文書名: 2.7.3 臨床的有効性 - 短腸症候群

25 / 135

Source: Module 5.3.5.1 CL0600-004 CSR, Table 9-5

Treatment Group	Response Category, n(%)				
	0	1	2	4	5 Off PS
Placebo	15 (93.8)	0	1 (6.3)	0	0
Tedaglutide 0.05 mg/kg/day	19 (54.3)	6 (17.1)	6 (17.1)	2 (5.7)	2 (5.7)
Tedaglutide 0.10 mg/kg/day	24 (75.0)	2 (6.3)	4 (12.5)	2 (6.3)	0
Total: (0.05 + 0.10 mg/kg/day)	43 (64.2)	8 (11.9)	10 (14.9)	4 (6.0)	2 (3.0)

PS量がベースラインから20%~100%減少した脱落者 (奏効スコア1~4) の数及び割合はテダグルチド 0.10 mg/kg/日群がプラセボ群に比べて大きかったが、合計スコア (total graded score) は 2.7.3-4)。スナップショットに有意差がない場合 (Week 20 及び

2.7.3 テダグルチド

Week 24 とともに奏効スコア 1~4 を示した脱落者) にテダグルチド 0.05 mg/kg/日群とプラセボ群との間で (奏効スコア全体を通じて) 臨床上有義のある数値差が認められたことから、さらに解析を進めることとした。追加解析の結果、テダグルチド 0.05 mg/kg/日群とプラセボ群との合計スコアの差の p 値 (rank-ANCOVA) は 0.007 であった。

Comparison*	Test Statistics	Degrees of Freedom	p-Value	p-Value (adjusted)*
0.10 mg/kg/day tedaglutide versus placebo	1.96	1	0.161	0.161
0.05 mg/kg/day tedaglutide versus placebo	7.32	1	0.007	
0.10 mg/kg/day tedaglutide versus 0.05 mg/kg/day tedaglutide	3.06	1	0.080	

飛び先プレビュー

文書名: 0000 CL0600-004: A Study of... 飛び先:

12 / 31

Module 5.3.5.1 CL0600-004 CSR, Tables and Figures

Table 9-5 あああああああああああああああ

適用 閉じる

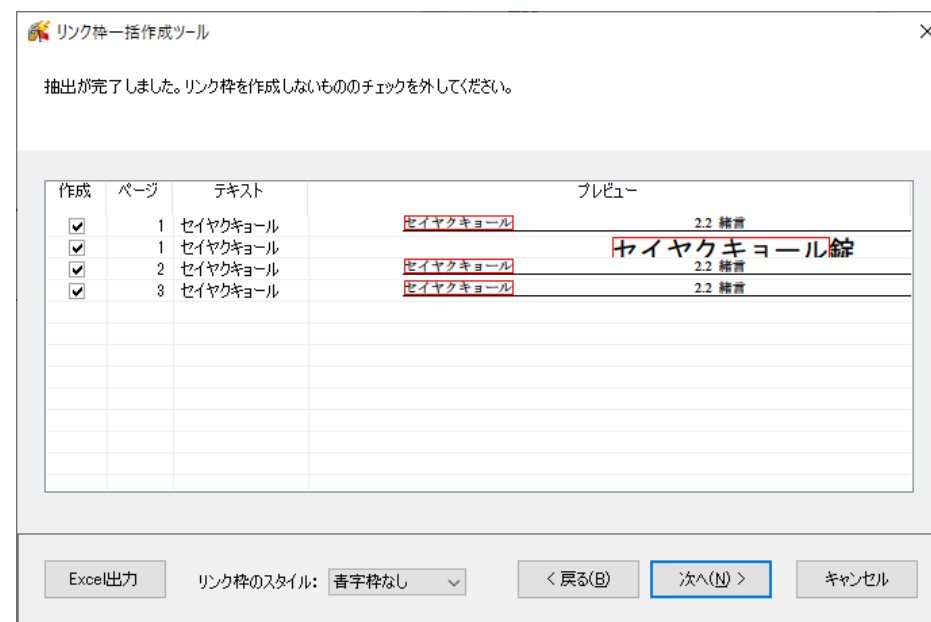
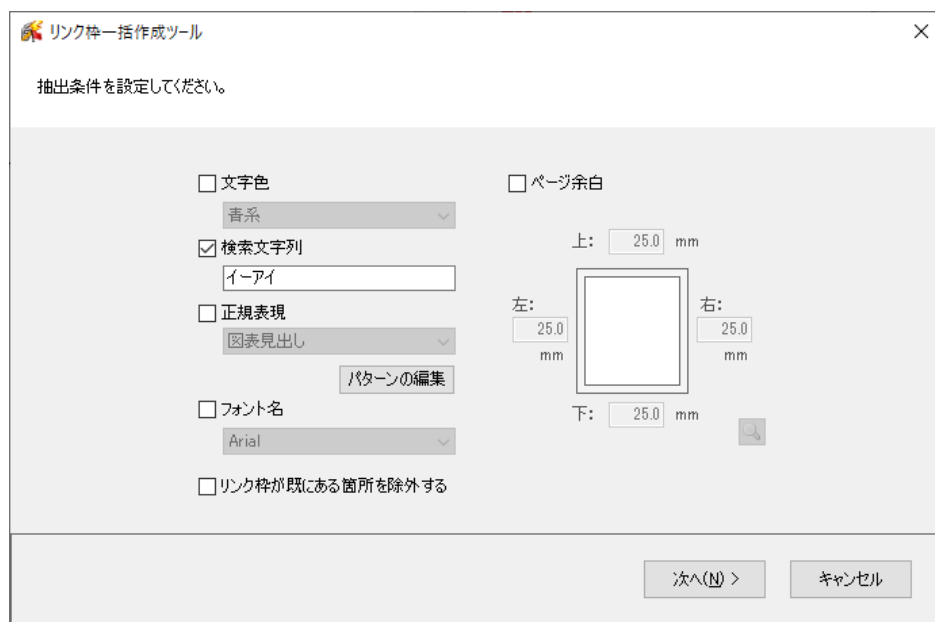
文書間リンクの一覧

リンク元
プレビュー
(編集可能)

飛び先
プレビュー
(編集可能)

2.1 リンク枠の作成

- リンク枠は、プレビュー上で文字列選択や範囲選択を使って自在に作成できます。
- 特定条件の文字列を検索して、まとめてリンク枠に変換する、**リンク枠一括作成機能**が搭載されます。



(参考：eCTDSupportersのリンク枠一括作成機能)

2. 2 飛び先の指定

飛び先プレビュー

文書名: 0000 CL0600-004: A Study of the Efficacy and Safety of Te... 飛び先:

ビュー: ツリー表示

シーケンス: 0000

(0, 841)

-3-3-3-intrinsic-factor-pk-study-reports
[5.3.3.3-1 試験番号 NSK-C-010 : XXX試験総括報告書](#)

①プルダウンから飛び先のリーフファイルを選択します

ああああああ

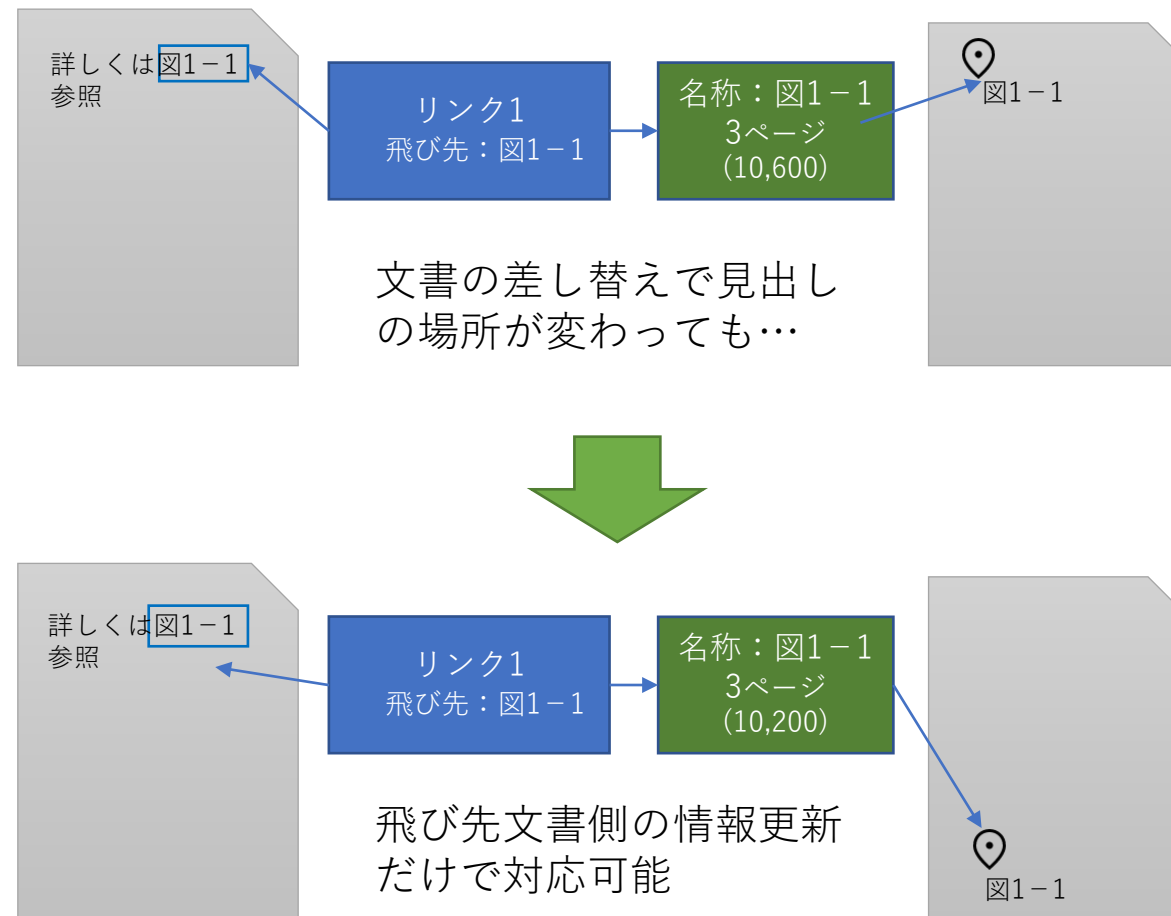
②あらかじめ登録しておいた飛び先候補を、プルダウンから選択します。

③適切な飛び先候補がない場合でも、ダブルクリックで任意の位置に指定できます。

-reports-of-human-pharmacodynamics-pd-studies
 -3-4-1-healthy-subject-pd-and-pk-pd-study-reports
[5.3.4.1-1 試験番号 NSK-C-013 : XXX試験総括報告書](#)
 -3-4-2-patient-pd-and-pk-pd-study-reports
[5.3.4.2-1 試験番号 NSK-C-014 : XXX試験総括報告書](#)
 i-reports-of-efficacy-and-safety-studies [indication:Short Bowe
 -3-5-1-study-reports-of-controlled-clinical-studies-pertinent-t
[CL0600-004: A Study of the Efficacy and Safety of Teduglutide](#)
[CL0600-004: Appendix 16.2.8 Individual Laboratory Measuremer](#)

飛び先候補について

- JetLinkでは、文書間リンクの飛び先を「飛び先候補」から指定します。
- 飛び先候補とは、PDF上の位置情報（ページおよび座標）に**名前**を持たせたものです。
- 利用イメージとしては**PDFの「移動先名（Named destination）」**に似ています。
- 飛び先候補は、リーフファイルのしおりやリンクからインポートできるほか、任意に作成・登録することができます。



3. 飛び先候補編集画面

飛び先候補の編集

新規 削除 # キーワードの編集 参照元の表示 インポート

飛び先候補の一覧

No.	ページ	名称	キーワード
1	1	2.7.3 臨床的有効性	
2	1	目次	
3	3	表目次	
4	5	図目次	
5	7	略号一覧	
6	9	要旨	
7	12	1 背景及び概観	
8	12	1.1 有効性試験の...	
9	12	1.1.1 成人SBS 患...	
10	12		
11	13		
12	13		
13	16		
14	20	2 個々の試験結果...	
15	20	2.1 臨床的有効性...	
16	20	2.1.1 成人SBS 患...	
17	43	2.1.2 小児SBS 患...	
18	65	3 全試験を通しての...	
19	65	3.1 成人SBS 患者...	
20	66	3.1.1 小児SBS 患...	

飛び先のプレビュー

2.7.3 臨床的有効性
テデュグルチド

1 背景及び概観

テデュグルチドは遠位回腸のL細胞によって分泌される天然型GLP-2のアナログである。PSは経静脈栄養[PN]及び経静脈栄養/補液[PN/IV]と同義であり、これらの用語は出典の表、図及び一覧表においても互換的に使用する。

本概要に記載されている内容、SHP633-305試験の中間解析及びSHP633-301試験の...

1.1 有効性試験の記述

本概要に含まれるSBS患者を対象としたテデュグルチドのプラセボ対照試験及び非盲検試験を表2.7.3-25に示す。

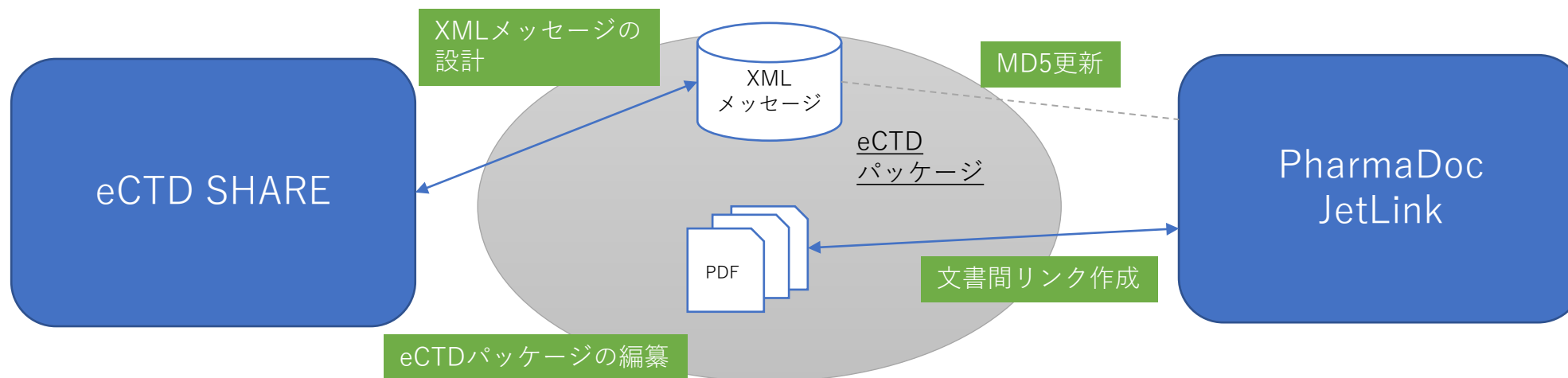
1.1.1 成人SBS患者を対象とした試験

成人SBS患者の治療におけるテデュグルチド0.05mg/kg/日の有効性の根拠は、海外で実施された第2相プラセボ対照試験であるCT0200020試験、CT0200021試験の試験及びその継続性...

適用 閉じる

他のシステムとの連携

- eCTD作成支援システム「eCTD SHARE」と組み合わせることにより、eCTDの編纂とPDFファイルのSR化を協調的に進めることができます。



評価版の提供予定について

- PharmaDoc JetLinkの製品評価版は、**2023年10月より提供予定**です。
- 評価をご希望のお客様は、弊社まで直接お問い合わせいただくか、本ユーザー会のアンケートに「**評価版希望**」とお書き添えください。
提供可能な状況になりましたら、弊社からご連絡いたします。
- たくさんのお客様のご応募を心よりお待ちしております。

**ご不明点等ございましたら、
お気軽にお問い合わせください**

ご清聴ありがとうございました

End of File