

DAGRI Data format: 農業経営データ解析タスクのための統一フォーマット構築

小林 暁雄¹ 大友 将宏¹ 石原 潤一¹ 馬場 研太¹

桂樹 哲雄¹ 森岡 幹² 坂地 泰紀³ 木村 泰知⁴

¹ 農研機構 農業情報研究センター² 株式会社 AIREV

³ 北海道大学⁴ 小樽商科大学

kobayashi.akio544@naro.go.jp

概要

NTCIR-19 DAGRI タスクでは、各都道府県で農業従事者に対して指導を行う農業普及指導員が作成を行っている、農業の年間経営計画および経営収支目標などを定めた農業経営類型及びその基礎データである経営指標について、公開されている PDF ファイルから情報抽出を行うタスクと、抽出されたデータを基に経営に関する質問に回答を行う質問応答タスクを実施する。本稿では、これら二つのタスクを円滑に行うため、抽出すべき情報の種類を定めるとともに質問応答で利用可能な機械可読な形式として、経営類型及び経営指標統一フォーマットの設計について解説を行う。

1 はじめに

我が国の農業従事者の平均年齢は 68.7 歳と高齢であり、農業人口は今後 20 年で 30 万人まで減少することが想定されている。このような状況下で国内の食料生産を維持するためには、ロボットや AI 技術を活用したスマート農業の拡充が喫緊の課題である。このような背景から、農業者に指導を行う農業普及指導員（以降、普及指導員）もこうした技術に精通する必要がある。しかし、普及指導員従事者も高齢化が進んでおり、若手の採用も積極的に進められる一方で、熟練の普及指導員の知識継承が課題となっている。これに対し、内閣府が主導する「研究開発と Society 5.0 との橋渡しプログラム¹⁾」のうち、農林水産省が令和 5~7 年度に進める「AI 農業社会実装プロジェクト」（以降、BRIDGE 農業 AI）では、普及指導員支援のための大規模言語モデル（LLM）の構築に取り組んでいる。この LLM は、スマート農業のような最新技術や熟練指導員の作成した文書

などから、普及指導員が必要とする知識を助言することを想定している。ここで対象としている普及指導員の業務の一つとして、農業経営体の営農状況の把握及び営農計画の策定支援がある。LLM がこれらを適切に補助するには、地域・規模・品目などの経営類型別の農業経営指標（以降、経営指標）などのデータの入手と読解が必要である。これらの資料はその多くが都道府県ごとに策定されており、一部はインターネット上で閲覧可能である。一方で、これらの資料は公開者ごとに独自に作成された PDF ファイル形式となっており、現行の LLM では、これらの文書を正確に読み解くことは困難である。このため、BRIDGE 農業 AI プロジェクトにおいては、経営指標の読解を特に重要なタスクとして位置付け、以下の課題に取り組んでいる。

1. ファイルから機械判読可能なデータへの変換
2. 経営指標を構成する表の関係性理解
3. 自治体で異なる書式に対する同定処理

そこで、NTCIR-19 DAGRI (Data Analytics for aGRicultural Information)²⁾を開催し、これらの課題をサブタスクとして定義し、国内外の研究者から解決策を募る取り組みを進めている。DAGRI を構成するサブタスクは以下のとおりである（詳細は [1] を参照）。

Subtask 1 経営指標ドキュメントからの経営指標データ抽出

Subtask 2 経営指標に関する QA

Subtask 3 年間農作業情報（栽培ごよみ）の画像データ読み取り

Subtask 1 と Subtask 2 については、経営指標データに関する一連の読解タスクを処理ごとに分割した

2) 国立情報学研究所 (2025) https://research.nii.ac.jp/ntcir/ntcir-19/NTCIR19_Kickoff_slides/DAGRI.pdf

1) <https://www8.cao.go.jp/cstp/bridge/index.html>

ものである。Subtask 1で、統一されていない公開者別の経営指標ドキュメントを統一されたフォーマットに落とし込み、Subtask 2で統一フォーマットに記載されている内容を読み解き、営農にかかわる様々な質問に回答するという、IE タスクと QA タスクで構成されている。

本稿では、これら2つの DAGRI サブタスクを設計するにあたり、公開者ごとに独自に作成されている経営指標を統一的に扱えるようにするための、統一フォーマットの定義とその設計について解説を行う。

2 関連研究

営農に関する経営指標については、[2, 3, 4]など、元々簿記などで現状の営農実態を把握することだけを目的としていた営農に関するデータ整理から、経営目標を設定し、それを実現するための正確な営農シミュレーションを行うためのデータ整備や評価システムの開発が行われてきた。また、近年では、LLM などを用いて、データ整備や営農シミュレーションを試みた研究も行われている [5, 6]。

これらのツールやシミュレーションツール、評価システムなどの整備が進んでいる一方、各自治体の普及指導員により構築されている経営指標や経営類型といった文書は、自治体がそれぞれ独自に構築しているエクセルファイルなどを変換した PDF ドキュメントで提供されている。このように、これらの文書はあくまで農業者や普及指導員自身が読むことが主目的となっており、営農シミュレーションなどに利用することは容易ではない。DAGRI タスクではこの問題を解決し、人手で作成された文書を入力として、高精度な営農シミュレーションが実行できるシステムの開発を目指す。

経営指標の表構造解析に関しては、[7]にて表解析のベンチマークデータセットが構築されている。このデータセットは長崎県の経営指標に限定されたものであるため、本稿では、これを拡張し他の都道府県などが公開している経営指標にも適用可能な統一フォーマットを定義する。

3 経営指標と経営類型

経営指標は営農における作付け計画を実行するために必要となる資源（土地、労働、資本、機械・施設、資材等）の種類・量と想定される収益・費用の推定値である。経営類型は経営指標に基づいて模範

表 1 経営指標と経営類型の主要項目

経営指標	
作業技術一覧	作物別の年間に必要な作業・機材・資材・労働力
作業時期	作物別の年間に必要な作業の作業時期
経営収支	作物別の収益・費用
機材一覧	作物別に必要な機材の一覧と取得費用・減価償却費用
経営類型	
収支目標	経営収支の目標額
経営規模	目標を達成するための労働力や栽培面積など
機材・資材	営農モデルに必要な機材と資材の一覧
農業種別	有機農業など、収支とは独立した営農目的や手段の種別

的な経営目標を仮定し、その目標を達成できる営農計画の具体像を設定するものである [2]。具体的には、収支計画、作付け計画、機械・施設投資計画、経営資源の外部調達計画などから構成される。より簡潔に表現すると、特定作物について、生産から販売までに必要な作業項目について、機材や労働力・費用および販売収益などを個別の表でまとめたものが経営指標であり、それらに基づいて、仮定の農業者（労働力と耕地面積）と経営目標を設定し、経営指標の作物と栽培面積の組み合わせを経営目標を達成できる形でモデル化したものが経営類型である。具体例として、付録に経営指標と経営類型文書の一部抜粋を掲載する（図 1）。経営指標側では収支が 10a 単位で計算されているのに対し、経営類型ではそれぞれ作付面積ごとに算出した値が組み合わさった合計値が算出されており、各合計値の値がこのモデルで目指すべき収支額となる。

各自治体や事業者が公開している経営指標・経営類型文書では、おおむね表 1 のような情報が記載されている。

これらの文書はそれぞれの作物や家畜などの担当者（多くの場合は普及指導員）が作成しているため、公開者間で共通したフォーマットが無いだけでなく、同一公開者内でも文書によって記載の形式が異なることも多い。

このような背景から、DAGRI タスクでは、自治体ごとの文書を比較しつつ、上記の項目に沿うような項目を持つ統一フォーマットを構築した。

3.1 統一フォーマット

DAGRI Subtask 1 のデータ抽出項目のためのスキーマであり、Subtask 2 における QA の知識源としてデータを一貫して取り扱うための、DAGRI 統一フォーマットの概要について解説する。フォーマットを構築するにあたり、公開者ごとの経営指標または経営類型文書の調査と比較を行った。調査対象

は、公開者単位とし、各サイト内またはそこからリンクされた文書のうち、経営指標または経営類型およびその両方が記載されているものを取得した。

取得された文書に対し、上述の経営指標と経営類型における各項目がそれぞれ含まれているか否か、また公開されている文書の数量と形式などについても調査を行った。

その結果を表 2 に示す。表中における最右項目の備考欄では、紙をスキャンしたものである場合には「スキャン画像」、各都道府県が作成している、「農業経営基盤の強化の促進に関する基本方針」と呼ばれる、自治体ごとの農業の特徴や目標の紹介とそれに基づく経営類型をまとめた文書から取得された場合には「経営強化方針」、南石らの構築したデータ形式 [3] に基づく場合は「FAPS-DB 形式」と記載している。

表より、すべての公開者で記載が必須となっている項目は存在していないことが分かった。また、調査項目に設定したすべての経営指標・経営類型項目を網羅している公開者も存在していなかった。このことから、統一フォーマットを作成するにあたっては、調査項目に対応するような網羅的なフォーマットを構築しつつも、いずれの項目も必須項目とすべきでないことが分かった。特に、「農業経営基盤の強化の促進に関する基本方針」文書については、もともと各都道府県の農業改良普及事業担当部署が事業紹介用に作成しているため³⁾、経営類型や経営指標を取りまとめる目的とは異なる文書である。このため、記載されている内容は各地で取り組まれている営農モデルの紹介としての側面が強く、含まれているデータは、簡単な経営類型にとどまっている。このため、経営指標のデータは得られないものの、各都道府県に作成が義務付けられていることから、表に示していない他の都道府県も文書を提供しており、経営類型のデータ数を確保する上で有用な文書である。

また、岩手県の文書については、FAPS-DB に収録されたデータ形式となっており、もともと営農シミュレーションに使用しやすい形式になっている。一方で、他の公開者とかなり異なる設定項目が多く、また営農に関する詳細な知識を必要とするため、参加者の負担も考慮し、単純に FAPS-DB の形式を統一フォーマットに採用することは困難である

3) 農業経営基盤強化促進法(昭和55年法律第65号)5条にて、定期的な作成が義務付けられている。 https://www.maff.go.jp/j/keiei/koukai/attach/pdf/chiiki_keikaku-150.pdf

と判断した。

これらの調査結果を受けて、次のように統一フォーマットの主要項目を定義した。

前提表 経営類型における収支目標額、労働力、耕地面積（所有する圃場などの一覧）

栽培規模 各経営指標の作物と対応付けられる栽培面積を定義

経営収支（モデル全体） 収支目標額の算出根拠となる、収益・費用に関する諸項目を定義

資本装備と減価償却費 経営類型に必要な機材とその取得費用・年間償却額を定義

経営収支（作物別） 各作物単位の 10a での収益・費用の想定額

作業技術一覧（作物別） 各作物単位の作業技術一覧。10a 単位での作業時間や使用資材なども定義。

作業時期（作物別） 各作物単位の 10a 単位での年間の月（上中下旬）別の作業時間一覧表。

統一フォーマットでは、前提表～資本装備と減価償却費までの 4 項目が経営類型、作物別の 3 項目が経営指標と対応する。機材に関しては、表 2 に示す通り、経営指標・経営類型どちらの機材に関する項目も記載がある公開者は広島県だけとなっており、また、その広島県も経営類型側の表が一つあるだけで、その中にどの経営指標で使用するのかの注意書きが記載されている形式となっていた。このため、Subtask 1 の情報抽出タスクにおいては、いずれの公開者のデータを提供する場合でも、単一の表からの情報抽出となるため、広島県の形式に則り単一の項目とした。

これらの項目について、それぞれ構造化されたサブ項目を記載可能な JSON 及び Excel 形式でのフォーマットを構築した。付録の図 2 に、図 1 の北海道の経営収支より作成した統一フォーマットを示す。

経営収支については、経営類型・経営指標ともに同一のフォーマットとなっており、農業簿記（一般的な簿記に対し、仕訳が営農に特化したものとなっている）に準ずる勘定科目をサブ項目として設定した。

作物別の各フォーマットについては、サブ項目名の表記ゆれはあるものの、形式が公開者間でほぼ同一であった。一方で、経営類型については、それぞれ独自の内容も含まれており、統一フォーマットに

表2 公開者別経営指標・経営類型文書の記載内容比較

公開者	経営類型				経営収支	経営指標			データ数	データ概要	
	収支目標	経営規模	機材・資材	農業種別		作業技術一覧	作業時期	機材一覧		形式	備考
長崎県	○	○	○		×	○	○	○	119	PDF	
有参協 ⁴⁾	×	×	×	有機	○	×	○	△ ⁵⁾	537	PDF	
千葉県	○	○	○		×	×	×	×	1	PDF	経営強化方針
香川県	×	○	×		○	○	○	○	152	PDF	
熊本県	×	○	×	スマート	○	○	○	○	10	PDF	
鳥取県	×	○	×		○	○	○	○	107	PDF	
岩手県	×	×	×		○	○	○	○	126	Excel	FAPS-DB 形式
山口県	×	×	×		○	○	○	○	25	PDF	
広島県	○	○	○		○	○	○	△	8	PDF	スキャン画像
北海道	○	○	×	有機	○	○	○	×	3	PDF	スキャン画像
兵庫県	×	○	○		×	×	×	×	1	PDF	経営強化方針
JA 山形	×	×	×		○	×	○	○	1	PDF	

採用するか判断が難しいサブ項目も存在した。このようなサブ項目に対しては、対象とした公開者の半数以上で採用されているか否かで採用の可否を決定した。このため、特に FAPS-DB 形式における独自の項目や、経営類型の適用地域など、詳細な営農シミュレーションにおいては重要と思われる項目についても、殆どの公開者で記載がなかったため採用を断念した。

3.2 サンプルデータ構築状況

現在、DAGRI で使用するサンプルデータおよび評価データのための、統一フォーマットでのデータ構築作業を進めている。表2中の公開者ごとに並行して、一定数の経営類型または経営指標を対象として作業を行っている。各データの特性に合わせて、機械処理可能なものについてはそれぞれ処理を行うためのプログラムを作成し、必要に応じて人手で修正を行っている。特に、長崎県については、すでに人手による CSV 形式への変換を完了 [8] しているため、機械的に変換を進めている。一方で、人手以外では困難な文書については、すべて人手により統一フォーマット構築を行っている。特に、北海道と広島県の2者については、印刷物をスキャンした画像で構成された PDF ファイル形式であることや、複数種類の比較などが文書中に含まれているため（付録の図1）、人手により作業を行っている。また、印刷スキャンでない場合の文書であっても、PDF 化の際にテーブル構造などのメタデータは失われているため、OCR を実施する場合もある。このように、公開者ごとに記述形式が異なっているため、作業マニュアルも並行して公開者ごとに作成を進めている。以上のことから、公開者ごとにサンプルを準備するコストが大きいため、一部の公開者のみを対象としている。現在の構築状況を表3に示す。

表3 統一フォーマットデータ構築状況

公開者	構築件数	全体件数	作業方針
長崎県	2	119	機械処理のみ
北海道	3	3	人手のみ
広島県	6	62	人手のみ
鳥取県	1	107	人手+機械処理

マニュアルについては、DAGRI 参加者にも提供される抽出対象の詳細情報となるため、統一フォーマット本体よりも構築に人手を要する。このため、構築体制の見直しなどが今後の課題である。

4 おわりに

本稿では、NTCIR DAGRI タスクにおける、農業経営指標および経営類型に関する情報抽出タスクである Subtask 1、および、経営指標に基づく営農シミュレーションに関する QA タスクである Subtask 2 において共通して使用するデータを整備するための経営類型及び経営指標統一フォーマットの提案とサンプルデータ構築状況について報告した。統一フォーマットについては、都道府県や JA など 12 団体が公開している経営類型・経営指標文書を参考に、共通項目を設定することで構築を行った。サンプルデータ構築については、同公開者のうち、4 者の文書に対して構築を行い、合計 11 件構築が完了している。NTC では、サンプルデータの公開を 2026 年初頭を予定しており、公開タイミングまでにより多くの公開者を対象としてサンプルデータを追加で構築する予定である。

4) NPO 法人 有機農業参入促進協議会

5) △は記載はあるものの詳細な記述が無く、Subtask 2 で使用するには不十分と考えられる場合

謝辞

本研究は、内閣府研究開発と Society5.0 との橋渡しプログラム (BRIDGE)「AI 農業社会実装プロジェクト」JP23836805 の補助を受けて行った。

参考文献

- [1] 森岡幹. Ntcir-19 dagri. 言語処理学会第 32 回年次大会論文集. 言語処理学会, 2026.
- [2] 南石晃明, 前山薫, 本田茂広. 農業技術体系データベースと統合化された営農計画支援システム faps-db. 農業情報研究, Vol. 16, No. 2, pp. 66–80, 2007.
- [3] 南石晃明. 営農技術体系評価・計画システム faps の開発. 農業情報研究, Vol. 11, No. 2, pp. 141–160, 2002.
- [4] 大石亘, 松本浩一, 梅本雅, 東野裕広, 村岡賢一. 営農計画策定支援システム z-bfm の特徴と活用方法. 関東東海農業経営研究, Vol. 101, pp. 63–68, 2011.
- [5] 馬場研太, 佐藤正衛. 農業経営指標策定に向けた大規模言語モデル活用の可能性—農業文書の表形式への自動構造化実験—. 農業情報研究, Vol. 34, No. 2, pp. 51–67, 2025.
- [6] 馬場研太, 佐藤正衛. 大規模言語モデルを用いた数理計画問題の求解自動化実験—営農計画への適用—. 農業情報学会 2025 年度年次大会 講演要旨集, pp. 93–94, 2025.
- [7] 阿部瑞稀, 杉山陽菜乃, 中村彩乃, 前多陸玖, 坂口遥哉, 佐藤栄作, 木村泰知. PDF 形式の農業技術文書を用いた表構造認識ベンチマーク TOITA. 言語処理学会 第 31 回年次大会 発表論文集, 2025.
- [8] 杉山陽菜乃, 阿部瑞稀, 中村彩乃, 前多陸玖, 坂口遥哉, 佐藤栄作, 木村泰知, 小林暁雄, 大友将宏, 石原潤一, 桂樹哲雄, 川村隆浩. 農林業基準技術に含まれる表を対象とした pdf から csv へ変換する際の課題分析. 言語処理学会第 31 回年次大会論文集, pp. 3954–3959. 言語処理学会, 2025.

作物名: 大豆

1. 作業技術体系 (時間/10a)

項目	有機栽培	慣行栽培	備考
作型			
堆肥散布	0.06	0.09	
心土破砕	0.29	0.17	
耕起・整地・砕土	0.66	0.42	
施肥・は種	0.46	0.42	追肥を含む
雑草管理	5.61	3.40	有機は中耕・手取り除草、慣行は除草剤散布
病害虫防除		0.09	
収穫	0.44	0.84	慣行には収穫前の抜き草を含む
乾燥・調製			有機はほ場から乾燥施設へ搬入、慣行は共同乾燥施設へ搬入
出荷		0.05	
あと片付け	0.18		
	7.70	5.48	

注) 慣行栽培は北海道農業生産技術体系第5版から引用。

2. 経営収支 (円/10a)

項目	有機栽培	慣行栽培	備考
作型			
収入			
主産物収量 (kg)	210	260	
単価 (円/kg)	478.7	305.4	販売手数料差引価格注2)参照
副産物収量			
単価			
合計	100,524	79,404	
変動費			
肥料費	2,693	8,282	
種苗費	3,749	4,448	
農薬費		4,250	
諸材料費	1,700	1,027	生産資材、小農具、包装資材
動力光熱費	3,225	1,517	
賃料料金	7,482	1,304	有機は雑穀業者への共通経費委託、慣行は共同乾燥・調製施設への委託
計	18,849	20,828	
固定費			
農機減価償却費	38,610	29,735	
農機修理費	13,116	9,289	
計	51,726	39,024	
費用合計	70,575	59,852	
差引利益	29,949	19,552	
時間当たり差引利益	3,889	3,568	

注1) 慣行栽培の資料は前掲表による。但し、固定費はHokkaido営農Navi(東胆振版)を参考に試算。

注2) 主産物単価は、販売単価と直接支払交付金あわせて単価を算出している。

注3) 作物共済掛金及び運賃については、双方に計上していない。

経営体: 畑作・露地野菜

粗収益と経営費 (単位: 円)

作物名	大豆	秋まき小麦	春まき小麦	緑肥	にんじん	かぼちや	合計	補正
作付面積10a	4ha	3ha	0.5ha	0.5ha	1ha	3ha	12ha	
主産物	2,797,200	1,275,000	170,000	0	3,720,000	3,600,000	11,562,200	
交付金	1,391,600	1,033,500	137,800	0	0	0	2,562,900	
粗収益	4,188,800	2,308,500	307,800	0	3,720,000	3,600,000	14,125,100	14,125,100
経営費								
種苗費	149,960	197,730	32,325	42,300	353,290	471,600	1,247,205	
肥料費	107,720	340,500	67,575	12,240	211,600	360,000	1,099,635	
農薬費								
生産資材費						191,340	191,340	旬別労働時間
小農具費	40,000	15,000		2,500	1,140	5,820	64,460	の雇用費用
燃料費	129,000	111,630	15,260	16,815	60,300	318,900	651,905	(時間×870円)
作物共済費	35,440	16,620	2,250				54,310	1,502時間
雇用労賃					232,470	103,320	335,790	1,306,740
賃料料金								
水利費								
機械利用費	1,544,400	671,520	141,125	79,875	241,120	711,480	3,389,520	
機械修理費	524,640	184,920	20,660	19,010	139,830	227,970	1,117,030	
販売経費	528,720	132,450	17,660	0	489,600	288,000	1,456,430	費用補正
経営費合計	3,059,880	1,670,370	296,855	172,740	1,729,350	2,678,430	9,607,625	10,578,575
粗利益	1,128,920	638,130	10,945	-172,740	1,990,650	921,570	4,517,475	3,546,525

作物別収支合計 注) 旬別の雇用労賃を算入

図1 経営指標(上2表)と経営類型(下1表)の例; 北海道農業研究協議会「有機農業の経営指標(事例)」から抜粋

```
{
  "income": 400000c0,
  "income_unit": "円",
  "gross_revenue": 14125100,
  "gross_revenue_unit": "円",
  "sales_revenue": null,
  "sales_revenue_unit": null,
  "amount_of_yielding_main_product": null,
  "amount_of_yielding_main_product_unit": null,
  "unit_price_main_product": null,
  "unit_price_main_product_unit": null,
  "other_income": null,
  "other_income_unit": null,
  "management_cost_total": 10578575,
  "management_cost_total_unit": "円",
  "variable_cost_total": null,
  "variable_cost_total_unit": null,
  "seedling_cost": 1247205,
  "seedling_cost_unit": "円",
  "fertilizer_cost": 1099635,
  "fertilizer_cost_unit": "円",
  "pesticide_cost": null,
  "pesticide_cost_unit": null,
  "materials_cost": 255800,
  "materials_cost_unit": "円",
  "fuel_and_utilities_cost": 651905,
  "fuel_and_utilities_cost_unit": "円",
  "farm_tools_cost": null,
  "farm_tools_cost_unit": null,
  "insurance_cost": 54310,
  "insurance_cost_unit": "円",
  "packing_freight_fees": 1456430,
  "packing_freight_fees_unit": "円",
  "fixed_cost_total": null,
  "fixed_cost_total_unit": null,
  "machinery_depreciation_cost": 3389520,
  "machinery_depreciation_cost_unit": "円",
  "machinery_repair_cost": 1117030,
  "machinery_repair_cost_unit": "円",
  "facility_depreciation_cost": null,
  "facility_depreciation_cost_unit": null,
  "facility_repair_cost": null,
  "facility_repair_cost_unit": null,
  "labor_cost": 335790,
  "labor_cost_unit": "円",
  "land_improvement_and_water_cost": null,
  "land_improvement_and_water_cost_unit": null,
  "estimated_amount": null,
  "estimated_amount_unit": null,
  "imputed_labor_cost": null,
  "imputed_labor_cost_unit": null,
  "imputed_rent_paddy": null,
  "imputed_rent_paddy_unit": null,
  "imputed_rent_upland": null,
  "imputed_rent_upland_unit": null,
  "imputed_rent_orchard": null,
  "imputed_rent_orchard_unit": null,
  "imputed_rent_greenhouse": null,
  "imputed_rent_greenhouse_unit": null,
  "imputed_interest_on_capital": null,
  "imputed_interest_on_capital_unit": null,
  "production_cost_before_byproduct_deduction": null,
  "production_cost_before_byproduct_deduction_unit": null,
  "profit": null,
  "profit_unit": null
}
```

図2 経営収支の統一フォーマットサンプル