

قاعدة بيانات

قاعدة بيانات أرشيف البيانات والمواقف غير الرسمية - مصر 2011-2018

الدليل والمنهجية

مخزن البيانات الحرة

Open Data Tank Initiative e.V

أولاً: تقديم مختصر للدليل:

تضم قاعدة بيانات أرشيف البيانات غير الرسمية في مصر البيانات الصادرة من المجموعات غير الرسمية (الكيانات والأحزاب والحركات السياسية والاجتماعية ومؤسسات المجتمع المدني)، التي تتعلق بخسائر بشرية أو فعاليات أو صراع سياسي أو مجتمعي على خلفية سياسية أو اجتماعية أو طائفية أو رياضية أو إقليمية خلال الفترة بين (25 يناير 2011 حتي 31 ديسمبر 2018).

تعتبر قاعدة البيانات مفتوحة المصدر ومنشورة برخصة المشاع الإبداعي بحيث تشمل تفاصيل الوقائع التي حدثت خلال فترة 2011-2018، من البيانات وتفاصيل نوع البيان من "رفض وإدانة وأستنكار ومشاركة وتنديد وتضامن". من خلال قاعدة البيانات، تم إعداد إحصاءات وصفية بين عديد من المتغيرات (أبعاد زمنية وجغرافية للبيان، خلفية ونوع البيان، وتصنيف للحصر الشامل للبيانات "حصر شامل وحصر رسمي").

المنهجية التي تم الاعتماد عليها هي Methodology of Triangulation and Data Verification المعروفة في مجال العلوم الاجتماعية، وكانت مصادر المعلومات الأساسية هي المواقع الرسمية للكيانات عبر وسائل الاعلام، بحيث تمثل نسبة 70%، باعتباره المصدر الأكثر مصداقية من إجمالي الحالات، ثم مصادر اخري (بعض المواقع والصفح التي تهتم بنشر اخبار عن بيانات الكيانات) حيث تم الاعتماد عليها كمصادر تكميلية للمصادر الأساسية، وبلغت نسبتها 30.10%.

ثانياً: محتويات الإصدار:

- ملف قاعدة البيانات التفصيلي على هيئة xlsx والتي تحتوي على الإصدار المعلوماتي: نسخة كاملة تحتوي على قاعدة البيانات وورقة sheet لتمثيل البيانات إحصائياً.
- صور screenshots لجداول الإحصائيات.
- ملف pdf عن الدليل والمنهجية.
- ملف pdf عن ورقة تقييم الفجوات المعلوماتية.

ثالثاً: معايير إدراج واستبعاد الحالات/البيانات:

- 1 البعد الزمني : يشترط أن يكون البيان او الموقف قد صدر من الكيانات السياسية خلال فترة 2011 حتي 2018 .
- 2 البعد المكاني : يشترط صدور البيان داخل حدود جمهورية مصر العربية سواء في أماكن مفتوحة مثل الشوارع والميادين، أو أماكن مغلقة مثل الجامعات أو المساجد والكنائس أو أماكن الاحتجاز الرسمية.

رابعًا: معايير ووحدة التعداد الإحصائي:

أعداد البيانات الواردة في الجداول الإحصائية لا تمثل جميع الكيانات السياسية فقد تم اختيار بعض الكيانات السياسية ورصد وتوثيق البيانات والمواقف الصادرة منها .

- الأعداد الواردة في الجداول الإحصائية تعبر عن بيانات وأحداث لحظية وليست أحداث مستمرة.
- الغرض الأساسي من البناء المعلوماتي والإحصائي هو رسم الأنماط والسلوكيات العامة للبيانات الصادرة عن الكيانات. الأعداد داخل الجداول الإحصائية في البيانات ذات الخلفيات الأخرى غير السياسية قد تُمثل "نسبة" مما يحدث على أرض الواقع، ولكن يمكن اعتبارها أنها نفس النسبة "تقريباً" في مختلف النطاقات الزمانية والمكانية والظرفية، وذلك بسبب الاستعانة بمختلف المصادر على مستوى الجمهورية مع مراعاة عامل "المركزية الشديدة للمعلومات" عبر التنوع والتكيز على المصادر الغير مركزية والمحلية المختلفة.
- العدد صفر لا يعني عدم وجود بيانات مطلقاً أو واقعياً، ولكنها تعني عدم التوصل إلى بيانات حسب نفس آلية جمع المعلومات والمنهجية.
- العدد واحد بخصوص البيانات لا يعني إصدار بيان واحد، ولكنه افتراض لإصدار بيان علي الأقل .

خامسًا: آلية جمع المعلومات والمصادر:

* مصادر المعلومات:

- مصادر المعلومات التي تم الاستعانة بها في البناء المعلوماتي لقاعدة البيانات الأولية بشكل رئيسي هي "مصادر أولية" (معلومات تم التحقق منها لوجود المصدر الأساسي لنشأتها)، بينما تمت الاستعانة بالمصادر الثانوية كمصادر تكميلية لباقي التفاصيل.
- تم وضع رابط مباشر لتلك المصادر تفصيلياً حيث أنه لكل حالة في قاعدة البيانات .

يمكن تقسيم أنواع مصادر المعلومات كما يلي :-

1. المصادر الرسمية المنشورة عبر المواقع الرسمية للمجموعات : تمثل نسبة 70%، باعتباره المصدر الأكثر مصداقية من إجمالي الحالات
2. مصادر أخرى (صفحات تنشر بيانات المجموعات): حيث تم الاعتماد عليها كمصادر تكميلية للمصادر الأساسية وتمثل نسبتها 30.10%.

البيانات الوصفية أو الفوقية Metadata لكل بيان

- هي بيانات جديدة تصف البيانات الأساسية (مثل بناء متغير نوع الكيان وتقسيمها إلى فئات متعددة حركة سياسية و حزب سياسي و حمبة و رابطة رياضية و مركز حقوقي و أشخاص عامة).
- البيانات الوصفية هي أدق ما تم التوصل إليه من جميع المعلومات المُتاحة وفقاً لقواعد إحصائية في مجال العلوم الاجتماعية، ولكنها ليست بالضرورة كاملة الدقة (100% هي نسبة مطلقة غير واقعية في الأبحاث الاجتماعية) حيث أنه هناك معايير فرضيات واستنتاجات تم الاستعانة بها.

- تم تمييز أعمدة البيانات الوصفية داخل قاعدة البيانات بالخلفية ذات اللون الأصفر.

سادساً: منهجية التحقق من البيانات وعوامل تقييم المعلومات:

تم استخدام منهجية التحقق Methodology of Triangulation and Data Verification المعروفة في مجال العلوم الاجتماعية، والتي يتم فيها الاعتماد على مصادر مختلفة لنفس المعلومة أو مجموعة المعلومات، ويتم خلال تلك العملية تقييم المعلومات الواردة من مصادر مختلفة وفقاً لعدة عوامل للاستبعاد أو التعديل أو الدمج أو الإضافة.

- مراحل تفكيك المحتوى المعلوماتي والتعامل معه:

- 1. الفهم:** براءة وفهم السياق وطبيعة تركيب البيانات ومصدر المعلومات والمحتوى المعلوماتي الموجود كاملاً واستيعابه مع تجاوز المصطلحات المستخدمة أو أية أخطاء أو أغراض بشرية.
- 2. التقسيم والتصنيف:** بتفكيك وتحديد المحتوى المعلوماتي وتحديد وحدة التعداد الإحصائي وإدراج كل معلومة في خاناتها بقاعدة البيانات.
- 3. بناء تصنيفات للبيانات (البيانات الوصفية):** حيث يتم فهرسة البيانات وتكويدها بحيث تكون أدق ما تم التوصل إليه من جميع المعلومات المتاحة وفقاً لقواعد إحصائية في مجال العلوم الاجتماعية.
- 4. معالجة الفجوات المعلوماتية:** لأي من المتغيرات بإعادة بناء مستويات أخرى لمصادر المعلومات.
- 5. الاستنتاج:** باستنتاج معلومات جديدة صحيحة 100%، ولكنها غير موجودة في المحتوى المعلوماتي الحالي (مثل استنتاج نوع كل واقعة من خلال قراءة نص الخبر أو تحليل الواقعة).
- 6. الفرضية:** بافتراض معلومات بنسبة كبيرة قد تكون صحيحة، ولكنها غير موجودة في المحتوى المعلوماتي الحالي، وذلك وفقاً لمعايير مُحددة أكاديمياً وخبرات ورؤية عامة للباحث.

عوامل تقييم المعلومات:

- حسب نوع الواقعة ومكانها وزمانها.
- حسب العناصر المتداخلة في الواقعة.
- حسب كم التفاصيل المتوفرة للمصدر.
- حسب تطابق التفاصيل الإطارية لمعلومات المصدر مع الحقيقة.
- حسب كم التضارب والأخطاء بين التفاصيل الداخلية لمعلومات المصدر.
- حسب منطقية ومنهجية التفاصيل الداخلية لمعلومات المصدر.
- حسب تقادم المعلومات.

سابعًا: الإطار المفاهيمي وتصنيف البيانات:

تعريف مصطلح البيان او الموقف في قاعدة البيانات :

- مصطلح البيان او الموقف يمثل بيان او موقف صادر من كيان معين علي خلفية أحداث سياسية أو اجتماعية أو نقابية أو ثقافية أو نسوية أو دعوية أو رياضية .
- مصطلح " عدد البيانات " المستخدم في الجداول الإحصائية يمثل عدد البيانات التي تم اصدارها خلال فترة الرصد .

معايير تقسيم المعلومات:

يمكن توزيعها إلى أربع أنواع من المتغيرات كما يلي:

أ. متغيرات تتعلق بخلفية البيان:

* تم تقسيم الحالات وفقاً لخلفية البيان كما يلي:

1. سياسية: تمثل البيانات الصادرة حسب التغيرات السياسية في مصر خلال الفترة من 2011 حتي 2018 بمختلف أشكالها
2. اجتماعية : تمثل البيانات التي صدرت خارج إطار التغيرات السياسية مثل بيانات التضامن علي خلفية مطالب اجتماعية.
3. رياضية: مثل البيانات التي صدرت علي خلفية واقعة رياضية.
4. نسوي : تمثل البيانات التي صدرت علي خلفية نسوية مثل بيانات دعم المرأة.
5. ثقافية: تمثل البيانات المتصلة بوقائع ثقافية .
6. اقتصادية: تمثل البيانات ذات الخلفية الاقتصادية (مثل البيانات المتعلقة بقرارات صندوق النقد الدولي وقرارات زيادة الاسعار)
7. دعوية : تمثل البيانات التي صدرت علي خلفية أحداث طائفية .
8. نقابية : تمثل البيانات الصادرة عن النقابات والبيانات العمالية .

ب. متغيرات متعلقة بنوع البيان:

*تم تقسيم الحالات طبقاً لنوع البيان كما يلي:

- إدانة : يقصد بها قيام الكيان بأصدار بيان يدين فيه قرار ما أو تصرف ما صادر عن جهة معينة .
- تنديد : يقصد بها التنديد بفعل او تصرف معين صادر من جهات او اشخاص .
- استنكار : يقصد بها استنكار لفعل او تصرف صادر من جهات او اشخاص .
- رفض : يقصد بها رفض لفعل او تصرف او تحرك صادر من جهات او اشخاص .
- تضامن : يقصد بها التضامن مع كيانات اخري في موقف او بيان ما أو التضامن مع اشخاص أو كيانات في مطالبهم
- دعوي : يقصد بها البيانات التي تدعو إلي حدث ما أو فاعلية .
- مشاركة : يقصد بها مشاركة الجهة الصادر عنها البيان فيما اصدرته.
- غير معلوم : تشمل البيانات الغير معلوم نوعها .

ج. متغيرات تتعلق بالبيان:

* تم تقسيم البيانات وفقاً للمحافظة المتعلقة بأصدار البيان كما يلي:

- تم اعتبار التقسيم الإداري للجمهورية وهي 27 محافظة.

* تم تقسيم البيانات وفقاً للإقليم الجغرافي للبيان كما يلي:

1. المحافظات المركزية: تشمل محافظات القاهرة والجيزة والإسكندرية.
2. محافظات الدلتا: تشمل محافظات القليوبية والدقهلية والشرقية والغربية والمنوفية والبحيرة وكفر الشيخ ودمياط.
3. مدن القناة: تشمل محافظات بورسعيد والإسماعيلية والسويس.
4. محافظات الصعيد: تشمل محافظات الفيوم وبني سويف والمنيا وأسيوط وسوهاج وقنا والأقصر وأسوان.
5. المحافظات الحدودية: تشمل محافظة شمال وجنوب سيناء ومرسى مطروح والبحر الأحمر والوادي الجديد.
6. غير معلوم: تشمل الحالات الغير معلوم فيها المكان المتعلق بأصدار البيان.

- تم تقسيم البيانات وفقاً للسلطة المتصلة بالبيان كما يلي:

1. سلطة تشريعية .
2. سلطة تنفيذية .
3. سلطة قضائية.
4. غير معلوم.

- تم تقسيم البيانات وفقاً لنوع الجهات الموجه لها البيان كما يلي:

1. اشخاص او هيئات حكومية .
2. اشخاص او هيئات غير حكومية .
3. شخصيات عامة .
4. غير معلوم .

- تم تقسيم المطالب الصادرة في البيانات كما يلي :

1. مطالب سياسية
2. مطالب فتوية .
3. غير معلوم .



د. متغيرات تتعلق بالكيانات:

تم تقسيم الكيانات الصادر عنها البيان كما يلي:

- 1 - حزب سياسي.
- 2- حركة سياسية.
- 3- مركز حقوقي
- 4- رابطة رياضية .
- 5- شخصية عامة .
- 6- حملة .

تم تقسيم الكيانات الصادر عنها البيان كما يلي:

1. كيانات مستمرة .
2. كيانات غير مستمرة.

ثامناً: نظام الفهرسة والفرضيات:

- تم ترتيب الحالات/الصفوف بالكامل داخل قاعدة البيانات حسب النظام المُفهرس التالي للأعمدة:
 1. تاريخ البيان: مُرتبة حسب يوم صدور البيان.
 2. ثم المحافظة المتصلة بالبيان .
 3. ثم المكان المتصل بالبيان داخل دائرة قسم الشرطة (الحي أو القرية) Sub District Level إن وُجد.
 4. ثم اسم الجهة او الكيان الصادر منه البيان .

- تمت مراعاة أن تكون أسماء جميع الأشخاص في ملف الإكسل وفقاً للمعايير الأرشيفية للغة العربية، مما يُسهّل من عملية البحث بعد الضغط على **CTRL+F** أو إعادة استخدامها أكاديمياً بواسطة برامج إدارة البيانات، كما يلي:
 - ✓ الألف "أ" تُكتب "ا" بدون همزات.
 - ✓ التاء المربوطة "ة" تُكتب "هـ".
 - ✓ الألف المقصورة "ى" تُكتب "ي".

- تمت مراجعة قاعدة البيانات بالكامل، من أجل تقليل نسبة "الازدواج الحسابي" بتجنب التكرار، وذلك عن طريق عقد مقارنات تحليلية وواقعية بالنسبة للجهة الصادر عنها البيان وتاريخ البيان، من أجل الوصول إلى أكبر دقة ممكنة.

تاسعاً: معايير ورخصة نشر قاعدة البيانات:

- قاعدة البيانات مفتوحة بالكامل ومنشورة برخصة قواعد البيانات المفتوحة .Open Database License (ODbL) v1.0.
- تمت مراعاة مبدأي "عدم انتهاك الخصوصية" و "عدم جلب الضرر".
- تم نشر قاعدة البيانات بنفس البيانات الوصفية Meta data المستخدمة داخل ملف الإكسل دون حذف، من أجل المراجعات والتدقيق، مع إدراج المعادلات الإحصائية جميعها مفتوحة، كمزيد من الشفافية وإتاحة التعلم.
- تمت مراعاة الشفافية والدقة والالتزام الكامل بتحديد المنهجية وذكر المصادر لكل معلومة بشكل تفصيلي.

عاشراً: التحديات وإشكاليات المتعلقة بعمليات جمع المعلومات والأرشفة:

1. مركزية المعلومات بشكل عام في مصر.
2. صعوبة الوصول لمكان البيان.
3. عدم توفر المعلومات في بعض النطاقات، وندرتهما الكبيرة في بعض الحالات وذلك بسبب حجب عدد كبير من المواقع والصفحات الخاصة ببعض الكيانات .
4. المخاطر الأمنية.
5. عدم الخبرة القانونية الكافية وقلة استخدام المصطلحات القضائية الدقيقة لدى الجماعة الصحفية.
6. عدم الوعي المجتمعي أو الفئوي بأهمية وجدوى التوثيق.
7. التناقضات والمبالغات والتضليل في المعلومات.
8. عدم قدرة مصدر المعلومة على التعبير وإيصال المعلومات بسهولة.
9. الحفاظ على الخصوصية وعدم جلب الضرر.