

## Review of Native Format and Metadata Usage in Access to Information Responses

### Part I – Native Format

An assessment of native format usage requires consideration of the relevant legislation, best practices in Information Management (IM) and Access to Information and Protection (ATIPPA) and associated concerns. These areas, as reviewed by the Office of the Chief Information Officer (OCIO) and the Access to Information and Protection of Privacy (ATIPP) Office are considered below.

An understanding of the standard definition of “native” format as well as “near native” format as described in the information technology (IT) sector is also important as follows:

- Native file format refers to the default format in which a document was created and maintained by the creating software or system. (Reference: <https://www.metaspike.com/native-file-format-cloud-emails/>). It is noteworthy that most IT developers will create their own file formats that may be usable (at least initially) by their own software.
- Near-native format produces the data contained in the document with as much original information as possible in a different, often more easily accessible file type. An example of near native would be the conversion of a word file to pdf.

Some software has become widely used in the public and there has been recognition in the IT sector of the need for easy translation of documents into other formats for general practical use. For example, Microsoft Word's native file format is “.docx”, but users are given the option to translate/save in other formats like “.txt” or “.pdf” as long as the user has this, presumably properly licensed, software on their computer.

Dr. Anton Oleynik has put forward in his supplemental submission to the ATIPP Review that he would like a definition of native format to be added to ATIPPA as follows: *“a format that does not materially change the electronic information that was originally created, sent or received”*. In reviewing Dr. Oleynik’s proposed definition however, it appears more in line with the near native definition as identified above in that it is not seeking direct release of the original document but rather a document that does not ‘materially change’. Even in this definition there appears to be recognition of the practicality of this undertaking.

The production of native format documents provide proof of a record’s authenticity and integrity – as needed in some legal proceedings. The question that becomes raised is the practicality of providing records in this format on a regular basis and the potential concerns with the information content released.

Section 11.2(c) of the ATIPPA, 2015 provides that “...a request shall... indicate how and in what form the applicant would prefer to access the record”. Further, Section 13(1) provides “The head of a public body shall make every reasonable effort to assist an applicant in making a request and to respond without delay to an applicant in an open, accurate and complete manner.”

Taken together, the spirit of ATIPPA appears to be to provide and assist applicants with the records in as complete a manner as possible and in the format that is preferable to the applicant. There appears to be breadth in the legislation to allow other formats, not just native, as long as transparency and accuracy of the information can be maintained.

ATIPP Coordinators in Newfoundland and Labrador regularly provide responses in near-native formats (i.e. pdf). In limited cases, Coordinators will provide native format files and will do their best to assist in providing in the form requested by applicants. As noted in a recent (informal) survey of eight jurisdictions (please refer to Appendix A), most jurisdictions are similar, transferring their responses to pdf format for applicant ease of use and for redaction purposes. Several jurisdictions identify that they provide native formats when requested and practical.

Section 4.1 of the Management of information Act (MOIA) outlines how public bodies should manage electronic records throughout their lifecycle as follows: “A requirement under this Act to retain a record is satisfied by the retention of electronic information where (a) the electronic information is retained in the format in which it was made, sent or received or in a format that does not materially change the electronic information that was originally created, sent or received; and (b) the electronic information will be accessible, and capable of being retained for subsequent reference, if required, by a person who is entitled to have access to the information or who is authorized to require its production.” Clearly, the MOIA foresees a potential need to keep the information in a near-native format that does not materially change the content, and identifies the clear link to record access. The Act does not exclude the ability to publish records in a near-native format that protects the integrity of the records. In addition, one of the OCIO IM principles and certainly one used in industry best practice, identifies: “Providing information authenticity, integrity and security to protect information holdings from loss, inappropriate access or use, disclosure, alteration, removal or destruction; thereby ensuring confidentiality, integrity, availability and accountability over time”. (IM&P Principles). Clearly the intent is to ensure records are not inappropriately altered – it is therefore incumbent on public bodies to ensure records are not materially changed.

In addition to the ATIPPA, 2015 and the MOIA as well as common practice, the OCIO is aware of two related legal proceedings in relation to ATIPP responses in native format:

1. 2017 CanLII 70006 (ON IPC) | St. Michael's Hospital (Re) | CanLII
  - The applicant requested his personal health information in native format

- Hospital argued that extracting raw data in native format would require a “translating program or mechanism from the vendor” so as to make the data readable
  - Court determined that the Ontario legislation did not require the hospital to provide records in the native format where their current software provided access to the data through custom software queries (see p. 62)
2. 2020 CanLII 35641 (NL IPC) | Memorial University of Newfoundland (Re) | CanLII
- Applicant requested records in their native format
  - Applicant wished to view the electronic native format in order to review the metadata of the record to verify the authenticity of the record.
  - The Commissioner determined that the right of access to records does not routinely include a right to access them in their native format under ATIPPA, 2015.
  - Decision - Compliance with section 20(1)(b) and Records in their Native Format
    - Section 11(2)(c) of ATIPPA, 2015 states that “A request shall indicate how and in what form the Complainant would prefer to access the record.” This section does not place an obligation on the public body to provide the records in the exact form requested.
    - Memorial explained within three days of the request being received, on September 13, that they may not be able to comply with his request for records in their native format and why. In spite of this, Memorial was able to provide the records in their native format, .msg during the course of our investigation. Therefore, this Office finds that Memorial provided this even though it was not required by the Act and did not fail in its duty to assist the Complainant in this regard.

The following are concerns with providing information in their native format, some are the same as those raised by Memorial University in their ATIPPA Review submission:

#### Information Access and Protection:

- The OCIO operates in a comparable environment as Memorial University in that there are hundreds of systems propriety, custom built and off the shelf that are managed and maintained on behalf of Government. Presently, to comply with ATIPPA’s duty to assist, the applicant is provided with easy to use information. This information is converted into readable text (near native) which the applicant can understand and use for their purpose. Some issues with regular use of native format include:
  - Vendor owned software may require vendor support to access – which requires a contract and payment (if the vendor is still in business), and adds additional time to access information
  - legacy technology may not allow access to the underlying software to create reports and data needed - if accessible, it may require IT specialists to interpret and may still require transfer to a near native format for ATIPPA purposes
- Information provided in native formats may be manipulated by a user. This calls into question key information protection principle of integrity of the information provided. It is industry best practice to

provide applicants with records that cannot be easily manipulated (i.e. pdf) to protect integrity and authenticity.

Operational:

- Not all ATIPP Coordinators would have the necessary software and knowledge to operate the system to access the records. (i.e.: AutoCAD). This results in an additional cost for training and licenses.
- Providing records in native format could increase IT storage requirements (and cost) as native files are larger than the comparable near native format (i.e.: pdf). Near native format are usually compressed when converted, resulting in smaller file size. Over time, native format usage would lead to significantly higher costs to Government and the taxpayer and potentially will impact the time it takes to search for responsive records.
- Information provided will not be easily redacted using existing software, in fact, potentially the native proprietary software may not allow any redaction software unless printed and manually redacted (e.g. moving to near native format). This could lead to breaches in the mandatory provisions of ATIPPA, 2015 and add time to the process.

Administrative:

- To provide responsive records in the native format will require more time for Coordinators due to the reliance on an IT specialist to first access the record in a complex/proprietary system, review for redactions and then identify a method to redact (if possible) in the native format. These extra steps would affect the timelines of the response.

The ATIPP Office has identified the following concerns:

- providing documents in their native format will pose an additional administrative burden on public bodies when dealing with requests that involve multiple records. This burden would become more problematic the larger the request. For instance, once a Coordinator locates all responsive records they are generally combined into one document for efficiency. These are then reviewed to identify exceptions and redactions are made. Without the use of redaction software either an IT specialist or printing the documents may be required to apply redactions – these actions would ultimately result in a near native format.
- the ability to provide native format where redactions are not required may seem reasonable when the request involves a small number of records – for example one or two emails, or other ATIPP requests at the same time, it would be extremely administratively burdensome to consider disclosure in native format where there are dozens, hundreds or thousands of individual emails.
- By way of an example – consider the case of an ATIPP request for which several hundred emails were considered responsive and native format was required :
  - a Coordinator would not combine all emails into one document for review, rather each email would need to be kept in its original format

- the records would be difficult to organize, maintain and review which could lead to records being missed or overlooked
- there would be increased difficulty to complete consults, either internal or external:
  - Coordinators would have to keep track of the various emails and which were sent to whom
  - feedback would be difficult to track as individuals wouldn't be able to refer to a specific page number
  - for larger files, there would be no ability to compress records to reduce the size which could cause difficulties with sharing
- redaction would need to be determined for each document separately
- a method to release the documents in a native format would need to be determined
- the applicant would need to also have the specific licensed software to access the documents when released

## Part II - Metadata

Metadata release adds a complexity to ATIPP as well. Metadata provides information about one or more aspects of the content in a record. Some examples could include: means of creation of the data, purpose of the data, time and date of creation, creator or author of the data, location on a computer network where the data was created, standards used, file size, data quality, source of the data, process used to create the data. Not all records have each of these pieces of information and without context and a point of reference, it might be impossible to identify metadata just by looking at it; an IT specialist help may be necessary.

The argument that emails, when converted into pdf format, lose their metadata and are thus materially changed, is not entirely true. Converted files are provided in a near native format which is often more accessible for applicants. Near native format provides the following metadata in the pdf version of the email:

- Sender and receiver
- Subject line
- Date and time email was sent
- Time email was received

There are concerns with releasing metadata including:

IT Security:

- There is an increased potential for cyber-attacks due to the inadvertent release of IT security information such as: Internet Protocol (IP) addresses, Server Names, infrastructure locations/details, etc.

- Disclosure of specific IT information in metadata may make IT assets, and the information they store/protect, more vulnerable to cyberattack
- If a cybercriminal were to access or manipulate personal and confidential government information found in metadata of a native file, the harm to citizens could be significant and harm to Government's reputation would be severe
- In the instance of emails in particular, the majority of additional information that would be provided in native format would need to be redacted due to security concerns. The metadata in the headers of emails include sensitive IT information including: IP addresses and servers

#### Privacy:

- There is an increased potential for data breaches due to the inadvertent release of personal information in the metadata

#### Administrative:

- Additional time will be required for Coordinators to contact the IT specialist for assistance, if needed and then analyze another layer of information (metadata) to determine the information to be redacted in accordance with permitted exemptions. These extra steps would affect the timelines of the response
- Complexity - the metadata adds another level to the validation and redaction process which will in turn affect ATIPP timelines
- Lack of Resources - many public bodies, small municipalities in particular, do not have dedicated IT resources who understand and can interpret the metadata and risks it poses

#### Summary

The OCIO and the ATIPP Office recognize that there may be instances where it would be reasonable to provide documents in native format (i.e. an applicant wants a few emails in their native format which don't require redactions), the software is easily accessed/understood, and disclosure of metadata would not pose a security or privacy concern. In these cases, public bodies should consider disclosure, in keeping with their legislative duty to assist.

Neither the ATIPPA, 2015, nor MOIA prevents this disclosure. Rather the ATIPP Act specifically requires Coordinators to provide assistance to an applicant, and applicants can seek information in a specific format.

However, the exceptions (therefore redaction) requirements under the Act and considerations around security, confidentiality, sensitive and personal information are significant as are the administrative impacts noted. Producing all responsive records in their native format could unreasonably interfere with operations of a public body and affect the timely access to the records requested. And in many cases

providing the record in native format may not provide any additional value to the applicant. Providing metadata, in particular, adds a significant risk to both Information Protection and Privacy.

Given the inherent risks, additional significant work required, potential limited additional value, and the current flexibility in the Act, is not necessary to add another clause, such as the proposed 20(x) to the legislation. The OCIO, along with the ATIPP Office, opposes the inclusion of a mandatory legislative requirement to provide records in native format, as it would create the potential for security and personal information breaches. Rather, the normal practice, to produce in a native format upon request is sufficient. Where records require redaction or metadata is requested, the Coordinator may need to discuss alternate approaches with the applicant.

The ATIPP office specifically notes that if the Review determines that the legislation should be amended to include reference to disclosure of records in their native format, consideration be given to making any such provision discretionary and based on reasonableness.

Appendix A  
Jurisdictional Review (2021)

Province	How Responsive Records Provided
Alberta	<ul style="list-style-type: none"> <li>pdf is the general practice for electronic disclosure - and is a secure way to redact and transmit record packages</li> <li>currently developing a policy instrument to standardize the process for the electronic release of responsive records</li> <li>instances where pdf not appropriate or the applicant has explicitly requested an alternate format the FOIP Office and client department will collaborate on best approach</li> <li>Open Government Portal - departments are encouraged to release digital publications in pdf format as it is considered a more open format than propriety word processing and email formats such as Microsoft Word or Outlook the docx format is used if the document contains a fillable form</li> </ul>
British Columbia	<ul style="list-style-type: none"> <li>Release records in native format if there is no severing that is required</li> <li>Files must be converted to pdf and disclosed in that format for our severing software to work</li> <li>Excel would be the most common native format used</li> </ul>
New Brunswick	<ul style="list-style-type: none"> <li>most departments disclose in electronic format converted to pdf to facilitate redactions</li> <li>electronic disclosure in original format is used in limited cases (e.g., upon request) and where no redactions are needed</li> </ul>
Nova Scotia	<ul style="list-style-type: none"> <li>do not release records in native format</li> <li>exceptions - database extracts (which use excel format which allows data to be manipulated, or picture/video files which don't require severing information)</li> </ul>
Prince Edward Island	<ul style="list-style-type: none"> <li>disclose records to applicants in the following formats:             <ul style="list-style-type: none"> <li>paper</li> <li>pdf - if disclosing as electronic records, usually copied to a USB</li> <li>excel - if the records originate in excel format</li> </ul> </li> </ul>
Saskatchewan	<ul style="list-style-type: none"> <li>most records disclosed as either paper or electronic copies converted to pdf format</li> <li>if a source format was requested it would be provided if no redactions were required</li> </ul>
Yukon	<ul style="list-style-type: none"> <li>similar approach to PEI - most requests are provided via PDF</li> </ul>
Canada	<ul style="list-style-type: none"> <li>federal legislation requires the "duty to assist" (e.g. provide access in the format requested), subject to the Regulations – in particular, when privacy, confidentiality, and security considerations would not be compromised, the record exists in that format, and it would not be unreasonable or impracticable</li> <li>most common practice for federal institutions is to provide redacted electronic documents in PDF format</li> </ul>