

Significant Significant Properties

R. van Veenendaal
National Archives of the Netherlands
The Hague
The Netherlands
remco.van.veenendaal@nationalearchief.nl

P.C.M. Lucker
National Archives of the Netherlands
The Hague
The Netherlands
pepijn.lucker@nationalearchief.nl

C.D. Sijtsma
University of Amsterdam
Amsterdam
The Netherlands
c.d.sijtsma@gmail.com

Conference Theme(s) Addressed: Sustainable digital preservation approaches and communities

Keywords: significant properties, digital preservation, community resource

ABSTRACT

The digital preservation community has clearly invested a lot of effort in the study of significant properties, especially in the first decade of this century. Significant properties have entered our standards, processes, tools and systems.

But: they have no single definition. They are not always measurable or automatically extractable. Their significance is in the eye of the stakeholder, and some stakeholders will raise their eyebrows questioning when you put a discussion about their significant properties on the agenda.

The above challenges and recent research indicate that there is still a need for the study of significant properties – more research needs to be done. We find it therefore somewhat contradictory to notice, that existing information on significant properties seems to be disappearing from the community’s radar, and that there seem to be fewer projects and initiatives that continue the study of significant properties in this second decade of this century.

At the National Archives of the Netherlands, we are collecting both theoretical and practical research on significant properties, and are working on a definition and database of so-called ‘significant significant properties’: those properties of information types that most preservation practitioners consider significant in most contexts. These properties are almost certainly significant to your stakeholders and can be likened to basic ingredients that absolutely need to be available in every kitchen, regardless of the meal that’s being prepared. The table salt of the preservation world, as it were. You may not even have to discuss them with your stakeholders. And then there are those properties that some practitioners consider significant in some contexts, or properties that you identify in your discussions with your stakeholder. The significance of these properties needs to be discussed with stakeholders, similarly to adding specific extra ingredients to the kitchen’s pantry after discovering your guests love Italian or Indian cuisine.

Our work is meant as a starting point for a shared, community-owned database with knowledge on significant properties, similar to the UK National Archives’ PRONOM technical registry and Yale University Library’s Wikidata initiative towards a digital preservation technical information data management system.

1 INTRODUCTION

This short paper is not about (re)defining the term significant properties. We decided to embrace the definition that is one of the most common in the international digital preservation community, and that works for us: “*The characteristics of an Information Object that must be maintained over time to ensure its continued access, use, and meaning, and its capacity to be accepted as evidence of what it purports to record.*” [1].

In the first decade of this century, there was a growing awareness of the need for digital preservation to maintain access to digital information and the role of significant properties in that process [2]. As a result, various projects and initiatives studied significant properties, proposed methodologies for identification and interpretation of significant properties and published significant properties of specific object / record / information types.

In this second decade of this century, the community’s focus seems to be more on applying the research results of the first decade. Tools for extracting properties and preservation planning were (and are being) improved (e.g. JHOVE¹ or PLATO²), and significant properties are found in most digital preservation systems (e.g. Archivemática³, Preservica⁴, RODA⁵ and Rosetta⁶). The place of significant properties in the digital preservation process was discussed. One should perhaps consider ‘preservation intent’ before discussing significant properties [3].

¹ JHOVE is a file format identification, validation and characterisation tool, see <http://jhove.openpreservation.org/>.

² PLATO is a tool for preservation planning, see <http://www.ifs.tuwien.ac.at/dp/plato/intro/>.

³ Archivemática is an open-source digital preservation system, see <https://www.archivematica.org/>.

⁴ Preservica is a commercial digital preservation system, see <https://preservica.com/>.

⁵ RODA is a digital repository system, see <https://www.keep.pt/en/productos/roda/>.

⁶ Rosetta is a digital asset management and preservation solution, see <http://www.exlibrisgroup.com/products/rosetta-digital-asset-management-and-preservation/>.

As a community, we have clearly invested a lot of effort in the study of significant properties. Bettavia [4] showed that relational and external significant properties do not fit within the OAIS reference model's entities, and is one example of the fact that there is still a need for the study of significant properties – more research needs to be done.

We find it therefore somewhat contradictory to notice, that existing information on significant properties seems to be disappearing from the community's radar, and that there seem to be fewer projects and initiatives that continue the study of significant properties in this second decade of this century. A warning sign is the fact that the often cited resources from www.significantproperties.org.uk have moved from a live website to the UK Web Archive⁷.

Even though we know there are still challenges w.r.t. significant properties, we also know that at some point we will have to address the issue of the significant properties of our Information Objects. This is preferably done with our stakeholders. Dappert and Farquhar explained that “significance is in the eye of the stakeholder” [5]. Which properties of their Information Objects do your Producers consider significant? Which properties are significant for your Consumers? Which properties can you as the organization responsible for the preservation of the Information Objects realistically maintain over time, given your policy, expertise, infrastructure and budget?

Discussing significant properties with your stakeholders is a challenge in itself. If we as a community aren't finished studying the subject, it is not so difficult to understand that many stakeholders will raise their eyebrows questioningly when you put a discussion about their significant properties on the agenda. You need to be prepared to explain what significant properties are, and how you are going to identify and deal with them. No matter how mature your significant properties methodology is, it would be very useful to be able to bring a reusable, community best practice-based list of potential significant properties per information type with you, instead of starting from scratch every time.

At the National Archives of the Netherlands (NANETH), we are working on such a list. We are collecting both theoretical and practical research on significant properties, and are working on a definition and database of so-called ‘significant significant properties’: those properties of information types that most preservation practitioners consider significant in most contexts. These properties are almost certainly significant for your stakeholders and can be likened to basic ingredients that absolutely need to be available in every kitchen, regardless of the meal that's being prepared. The table salt of the preservation world, as it were. And then there are those properties that some practitioners consider significant in some contexts, or new properties that you identify in your discussions with your

⁷ See

<https://www.webarchive.org.uk/wayback/archive/20130423072041/http://www.significantproperties.org.uk/>.

stakeholder. The significance of these properties needs to be discussed with stakeholders, similarly to adding specific extra ingredients to the kitchen's pantry after discovering your guests love Italian or Indian cuisine.

In section 2, we present a brief overview of the giants whose shoulders we're standing on. In section 3 we explain our work on significant significant properties. Section 4 contains our plans for the (near) future and questions to the digital preservation community.

2 SIGNIFICANT PROPERTIES

Knight and Pennock [2] present an overview of “*projects that have made an important contribution to the development of our understanding of significant properties*” in the first decade of this century, and is an influential work itself. Projects, initiatives and parties mentioned are: CEDARS, Digital Preservation Testbed, National Archives of Australia, DELOS, PLANETS, JISC-funded significant properties projects and InSPECT. The contribution of the United States National Archives and Records Administration (NARA) to the study of significant properties (e.g. by their Essential Characteristics Team and Technical Analysis Team)⁸ must also not be forgotten. The SCAPE project (2011 – 2014) reiterated the importance of significant properties, e.g. by including them in their Catalogue of Preservation Policy Elements⁹. The 2012 version of the ISO 14721:2012 OAIS reference model defines Transformation Information Properties (TIP) as “[a]n Information Property the preservation of the value of which is regarded as being necessary but not sufficient to verify that any Non-Reversible Transformation has adequately preserved information content” in an attempt to introduce uniform significant properties terminology.

These projects, initiatives and parties are examples of the shoulders of giants that we are thankful to be able to stand on.

3 SIGNIFICANT SIGNIFICANT PROPERTIES

3.1 Significant Properties at NANETH

We published our preservation policy in 2015¹⁰. One of the challenges raised in this policy is how we want to identify and deal with significant properties. As part of meeting this challenge, we looked at what others were doing and collected and studied previous work on significant properties. We noticed the aforementioned shift of community focus – from theory to practice – in dealing with significant properties in published research and by talking to colleagues from other organizations. In

⁸ See <https://www.archives.gov/files/era/acera/pdf/significant-properties.pdf>

⁹ See <http://wiki.opf-labs.org/display/SP/Catalogue+of+Preservation+Policy+Elements>

¹⁰ See <https://www.nationaalarchief.nl/sites/default/files/field-file/National%20Archives%20of%20the%20Netherlands%20preservation%20policy.pdf>

the impact assessment parts of projects in which we connect government archives to our e-Depot¹¹, we also noticed that the topic of significant properties came up in discussions with these stakeholders – our Producers.

Rather than starting each discussion with a Producer from scratch, we wanted to be able to bring a reusable, community best practice-based list of (potential) significant properties with us. Enter the idea for significant significant properties.

We presented our initial ideas about significant significant properties on a Dutch significant properties event on the first International Digital Preservation Day¹². We also shared our ideas with the Open Preservation Foundation's Archive Interest Group¹³, where we work with the Danish and Estonian National Archives, Preservica and the Dutch National Library on significant properties for spreadsheets and suitable file format migration strategies for preserving those properties.

The feedback we received made us decide to also want to share our work on significant significant properties with the wider digital preservation community at iPres 2018.

3.2 Information types

The list of potential significant properties is huge. In order to subdivide our challenge in smaller parts, and for presentation purposes, we decided to list significant properties per object / record / information type – a common approach in significant properties reports. For this, we reused your list of information types from our list of preferred and accepted formats¹⁴:

- Audio
- Database
- Document (text-based documents)
- Email (messages)
- Image (raster graphics)
- Presentation
- Spreadsheet
- Vector image
- Video

In order to establish this list, we took a look at community best practices w.r.t. preferred and accepted formats. Examples are:

- Data accompanying [6]¹⁵
- United States National Archives and Records Administration¹⁶

¹¹ See <http://openpreservation.org/blog/2017/05/17/preservation-impact-assessments-how-preservation-tools-support-naneths-connection-projects/>.

¹² See <http://www.ncdd.nl/op-15-hoog-digitale-erfgoed-vieren/> (in Dutch).

¹³ See <http://openpreservation.org/knowledge/interest-groups/archives/>.

¹⁴ See <https://www.nationaalarchief.nl/sites/default/files/field-file/National%20Archives%20of%20the%20Netherlands%20preferred%20and%20acceptable%20formats.pdf>.

¹⁵ See

<https://www.ideals.illinois.edu/bitstream/handle/2142/47421/FileFormatStatistics.pdf?sequence=4>.

- State Archives of North Carolina¹⁷
- Library of Congress Digital Preservation Program¹⁸
- ArchiveTeam¹⁹
- 4TU Centre for Research Data²⁰
- Data Archiving and Networked Service²¹
- British Library Digital Preservation Team (shared at DPC wiki page)²²
- United Kingdom National Archives²³

3.3 Property categories

For an even more detailed subdivision, we reused the common categorization of significant properties of the InSPECT project:

- **Appearance:** features related to the visual presentation of information objects to certain users at the moment of interaction, such as font, size, layout, etc.
- **Behavior:** characteristics that make behavior of, interaction with, or functionality of the information object possible, such as hyperlinks, formulas, etc.
- **Context:** characteristics of the organizational, functional and operational environment at the time of creating, receiving, storing and / or using information objects, and any relations of the information object with other information.
- **Content:** characteristics of the content of the information object, such as the text, images, recorded sound, etc.
- **Structure:** characteristics that specify how parts of the information object are organized and related to each other, such as embedded objects, paging, headings, etc.

3.4 Template and example table

By combining the information types and property categories, a template for significant significant properties, presented here as a table, emerged:

¹⁶ See <https://www.archives.gov/records-mgmt/policy/transfer-guidance-tables.html>.

¹⁷ See <https://archives.ncdr.gov/documents/file-format-guidelines-management-and-long-term-retention-electronic-records>.

¹⁸ See <http://www.digitalpreservation.gov/formats/fdd/descriptions.shtml>

¹⁹ See http://fileformats.archiveteam.org/wiki/Electronic_File_Formats

²⁰ See <http://researchdata.4tu.nl/en/publishing-research/data-description-and-formats/>.

²¹ See <https://dans.knaw.nl/en/deposit/information-about-depositing-data/before-depositing/file-formats>.

²² See http://wiki.dpconline.org/index.php?title=File_Formats_Assessments.

²³ See chapter 5 of

http://www.nationalarchives.gov.uk/aboutapps/fileformat/pdf/pronom_4_info_model.pdf.

Information type	Category	Properties	Significance
<TYPE>	Appearance		
	Behaviour		
	Context		
	Content		
	Structure		

Table 1- template for significant significant properties

We populate templates like Table 1 with information from our own experiences and community best practices, resulting in tables of significant properties per category, per information type, and with a ‘scoring model’ for significance: the more often a property is documented as a significant property in resources, the higher its significance and the more important it is to take into consideration and/or discuss with your stakeholders.

Because of differences in the way different organisations define and identify significant properties, the template and populated tables are over-generalized and provide little more detail than Information type, Category, Properties and Significance. When we gain more insight into the differences and similarities of the various approaches taken, we hope to be able to be more precise. As it is, the current state of affairs may not even prove to be a problem. The tables are meant as starting points for discussing significant properties with stakeholders. For detailed information and/or if further discussions with your stakeholders are required, you can follow the links in the Significance column to the original resource.

Even from the shortened example of the information type Email in Table 2 it can be gathered that context properties such as sender, recipient and subject could be considered significant significant properties, and should almost undoubtedly be preserved. The formatting of the email message body is another important property to talk about. In the same manner, all properties on the list can be discussed, e.g. in order of their relative significance. The results can be used in your significant properties methodology. If you discover properties that are not yet on the list of significant significant properties, you can add them for later reuse.

It is possible, that significant properties studies reused or will reuse existing studies, and that in some cases, they also reiterate certain significant properties. We see this as a strength: the follow-up studies confirm the significance of those properties. In our kitchen analogy: cook books repeat the message to can use salt or pepper to flavour your meals. They confirm the significance of those ingredients.

Our significant significant properties work is meant as a starting point for a shared, community-owned database with knowledge on significant properties, similar to the UK National Archives’ PRONOM technical registry²⁴ or Yale University Library’s Wikidata for Digital Preservation initiative²⁵.

²⁴ See <https://www.nationalarchives.gov.uk/PRONOM/Default.aspx>.

²⁵ See <https://guides.library.yale.edu/WikidataDigitalPreservation>.

Information type	Category	Properties	Significance
Email	Appearance	The formatting (font, orientation, color, layout) of the message body may change from the original, as long as the meaning doesn’t change)	[Preservica Technical Registry ²⁶], [NARA ²⁷]
		The exact position of attachments may not be important, but mentioning their existence is	[Digital Preservation Testbed]
		(...)	
	Behaviour	It must be possible to open attachments in a suitable application	[Digital Preservation Testbed]
		Hyperlinks to websites	[Digital Preservation Testbed]
		(...)	
	Context	Sender	[Preservica Technical Registry], [InSPECT ²⁸], [Digital Preservation Testbed]
		Recipient(s)	[Preservica Technical Registry], [InSPECT], [Digital Preservation Testbed]
		Subject (topic of the message)	[Preservica Technical Registry], [InSPECT], [Digital Preservation Testbed]
		Series/Original Order: Is the email transferred in distinct folders or directories that reflect a clear operational structure? (...) Then this structure is an EC.	[NARA]
		(...)	
	Content	Message body and markup (see [InSPECT] for preservation scenarios)	[InSPECT], [Digital Preservation Testbed]
	Structure	Relationship (relationship agent has with message: creator / sender / recipient / etc.)	[InSPECT]
		Message-id (unique identifier found in received emails)	[InSPECT]
		Schema/linkage: (...) if a reference to an attachment is present in email, both the reference and the attachment must be preserved as essential.	[NARA]
		(...)	

Table 2 - significant significant properties for Email

²⁶ We included properties from this registry as it is part of our e-Depot, and that makes the properties (more) significant for us.

²⁷ See <https://www.archives.gov/files/era/acera/pdf/significant-properties.pdf>.

²⁸ See

<https://www.webarchive.org.uk/wayback/archive/20130423072325/http://www.significantproperties.org.uk/testingreports.html>.

4 FUTURE WORK

4.1 Publishing and collaboration

At the time of writing, we are populating the significant properties database with the properties from the resources that we collected. Once this initial work is done and a first database of significant properties has been compiled, we will publish our work a more detailed report than this short paper. At the same time, we want our work to be a shared, community-owned database. It might be more useful to publish the information online, and e.g. in a wiki²⁹. We kindly invite the digital preservation community to share ideas for the best suited publishing platform for this kind of information.

One important aspect of this publishing platform is that it should facilitate collaboration. One approach would be to keep full control of the list ourselves, and to invite the community to send us suggestions that we weigh and then include (or not). This approach is somewhat comparable to how The National Archives' PRONOM Technical Registry is managed. Although this approach has benefits, we think we prefer a more interactive Web 2.0 approach³⁰, but know that we will immediately be faced with a trust issue: which change(r)s to trust? We are therefore very interested in the progress of Yale University's Wikidata for Digital Preservation initiative, and especially how they will deal with this trust issue, e.g. by introducing "signed statements". As explained in an Open Preservation Foundation webinar³¹, this issue will be investigated in the Google Summer of Code 2018³².

5 CONCLUSION

The digital preservation community has clearly invested a lot of effort in the study of significant properties. Significant properties have entered our standards, processes, tools and systems.

We know that there are still challenges w.r.t. significant properties and that more research needs to be done, but we also know that at some point we will have to address the issue of the significant properties of our Information Objects, and discuss them with your stakeholders.

In this short paper, we have presented our work on a definition and database of significant properties that can be used as a starting point in discussions with your stakeholders.

Our work is meant as a starting point for a shared, community-owned database with knowledge on significant properties, similar to the UK National Archives' PRONOM technical registry and Yale University Library's Wikidata for Digital Preservation initiative.

We want to invite the digital preservation community to share:

- Feedback on the concept of significant properties,
- Resources with information about significant properties, and
- Ideas for suitable publishing and collaboration platforms.

REFERENCES

- [1] Grace, S., Knight, G. & Montague, L., *Final Report* (InSPECT, 2009). Retrieved from <https://www.webarchive.org.uk/wayback/archive/20130423072041/http://www.significantproperties.org.uk/> on April 15, 2018
- [2] Knight, G. & Pennock, M. (2009). Data Without Meaning: Establishing the Significant Properties of Digital Research. *International Journal of Digital Curation*, 4(1), 159-174, <https://doi.org/10.2218/ijdc.v4i1.86>
- [3] Webb, C., Pearson, D., Koerbin, P. (2013). 'Oh, you wanted us to preserve that?!' Statements of Preservation Intent for the National Library of Australia's Digital Collections. *D-Lib Magazine* January/February 2013, Vol. 19, Number 1/2, <https://doi.org/10.1045/january2013-webb>
- [4] Bettavia, R. (2016) Mapping Significance of Video Games in OAIS. *Proceedings of the 13th International Conference on Digital Preservation (iPres 2016)*. Swiss National Library, Bern.
- [5] Dappert, A. and Farquhar, A. (2009). Significance is in the eye of the stakeholder. *Proceedings of the 13th European conference on Research and advanced technology for digital libraries (ECDL'09)*. Springer-Verlag, Berlin, Heidelberg, 297-308.
- [6] Rimkus, K., Padilla, T., Popp, T. and Martin, G. (2014) Digital Preservation File Format Policies of ARL Member Libraries: An Analysis. *D-Lib Magazine* March/April 2014, Vol. 20, Number 3/4, <https://doi.org/10.1045/march2014-rimkus>

²⁹ See <https://en.wikipedia.org/wiki/Wiki>.

³⁰ See <http://www.oreilly.com/pub/a/web2/archive/what-is-web-20.html>.

³¹ See <http://openpreservation.org/event/how-we-created-a-digital-preservation-portal-for-the-wikidata-knowledge-base/>.

³² See <https://summerofcode.withgoogle.com/>.